The American Economic Review

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

P14417

- Published at George Banta Co., Inc., Menasha, Wisconsin.

- THE AMERICAN ECONOMIC REVIEW, including four quarterly numbers, the *Proceedings* of the annual meetings, and *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September and December.

- Membership dues of the Association are $20.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. Subscriptions by nonmembers are $20.00 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $4.00 each. Each order for copies of either publication must also include a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

- Correspondence relating to the *Paper and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address may be sent to the secretary, Rendigs Fels, 1313 21st Avenue, South, Nashville, Tennessee 37212. To be effective, notice of change of address must reach the secretary by the 1st of the month previous to the month of publication.

- Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

Officers

President

JAMES TOBIN
Yale University

President-Elect

JOHN KENNETH GALBRAITH
Harvard University

Vice-Presidents

JAMES M. BUCHANAN
Virginia Polytechnic Institute
FRANCO MODIGLIANI
Massachusetts Institute of Technology

Secretary-Treasurer and Editor of Proceedings

RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

MARY JEAN BOWMAN
University of Chicago
CHARLES L. SCHULTZE
Brookings Institution
ROBERT DORFMAN
Harvard University
ARNOLD C. HARBERGER
University of Chicago
ROBERT EISNER
Northwestern University
JOHN R. MEYER
Yale University

Ex Officio Members

WILLIAM FELLNER
Yale University
WASSILY LEONTIEF
Harvard University

Cwk- #02168 - 64- P014417

THE AMERICAN ECONOMIC REVIEW

June 1971

VOLUME LXI, NUMBER 3, PART I

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors
BARBARA R. BERGMANN
JAGDISH N. BHAGWATI
PHILLIP CAGAN
GREGORY C. CHOW
CARL F. CHRIST
C. E. FERGUSON
HARRY G. JOHNSON
DANIEL MCFADDEN
ALVIN L. MARTY
HERBERT MOHRING
MARC NERLOVE
G. WARREN NUTTER
EDMUND S. PHELPS
VERNON L. SMITH

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

• Copyright American Economic Association 1971.

Articles

Optimal Taxation and Public Production: II—Tax Rules
Peter A. Diamond and James A. Mirrlees 261

A Short-Term Econometric Model of Textile Industries *Roger LeRoy Miller* 279

Information and Frictional Unemployment
Reuben Gronau 290

The Efficient Allocation of Subsidies to College Students *Stephen A. Hoenack* 302

Discrimination by Waiting Time in Merit Goods
D. Nichols, E. Smolensky, and T. N. Tideman 312

Optimal Mechanisms for Income Transfer
Richard Zeckhauser 324

The Optimal Quantity of Money, Bonds, Commodity Inventories, and Capital
Edgar L. Feige and Michael Parkin 335

A Utility Theory of Representative Government
Edwin T. Haefele 350

International Trade and Capital Mobility
Ernest Nadel 368

Peasants, Procreation, and Pensions
Philip A. Neher 380

A Model of Soviet-Type Economic Planning
Michael Manove 390

Transactions Costs and the Demand for Money
Thomas R. Saving 407

Communications

Interest Rates and the Short-Run Consumption Function *Warren E. Weber* 421

Unemployment and Inflation: A Cross-Country Analysis of the Phillips Curve *David J. Smyth* 426

Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry: An Alternative Approach *Richard D. Portes* 430

Income Taxes and Incentives to Work: Some Additional Empirical Evidence *D. B. Fields and W. T. Stanbury* 435

Fiscal and Monetary Policy Reconsidered:

- Comment *Bent Hansen* 444
- Comment *John H. Hotson* 448
- Comment *Barbara Henneberry and James G. Witte* 452
- Comment *Keith M. Carlson* 454
- Reply *Robert Eisner* 458

More on an Empirical Definition of Money: Note *David T. Hulett* 462

Dependency Rates and Savings Rates:

- Comment *Kanhaya L. Gupta* 469
- Comment *Nassau A. Adams* 472
- Reply *Nathaniel H. Leff* 476

A Property of a Closed Linear Model of Production: Note *David Levhari* 481

Economics of Production from Natural Resources:

- Comment *Richard F. Fullenbaum, Ernest W. Carlson, and Frederick W. Bell* 483
- Reply *Vernon L. Smith* 488

Employment and Rural Wages in Egypt:

- A Reinterpretation *James A. Hanson* 492
- Reply *Bent Hansen* 500

Errata 509

Notes 511

Mexican Economic Policy in the Post–War Period: The Views of Mexican Economists

By Leopoldo Solís*

TABLE OF CONTENTS

INTRODUCTION

I. BACKGROUND 3
The Agricultural Sector 3
Industrial Development 5
Distribution of Income 6
The Financial System 7
Public Finance 8
Balance of Payments 10

II. THE AGRICULTURAL SECTOR 11
Land Tenure 11
Underemployment and Colonization 14
Agricultural Policy 14
Capital Formation 14
Commodities and Markets 15
Technology and Credit 16
General Analysis of Agriculture 16

III. THE INDUSTRIAL SECTOR 17
Background 17
Foreign Investment 18
Industrial Protectionism 22

IV. THE DISTRIBUTION OF INCOME 25
V. STABILIZATION POLICIES 28
Monetary Aspects 28
The Inflationary Period 29
The Structuralist Position 34
The Orthodox Position 38
The Period of Stability 40

VI. FISCAL POLICY 42
Tax Changes Up to 1955 42
The Period of Tax Change of 1962–1964 43
The Tax Reforms of 1965 45

VII. THE EXTERNAL SECTOR 47
Foreign Trade 47
Tourism 50
Latin American Economic Integration 50

VIII. ECONOMIC DEVELOPMENT 53
The Process of Development in Mexico 55

IX. ATTEMPTS AT ECONOMIC PLANNING 56

X. ECONOMIC POLICY, ECONOMISTS AND MEXICAN NATIONALISM 57

REFERENCES 64

Mexican Economic Policy in the Post–War Period: The Views of Mexican Economists

By LEOPOLDO SOLÍS*

Introduction

In Latin America, where inflation and economic stagnation seem to be the common rule, the case of Mexico is almost unique. In the past twenty-five years it has experienced sustained economic growth at a rate of more than 6 percent a year, almost double the rate of population growth. Output per capita has reached 560 dollars and Mexico ranks twelfth among the economies of the free world in terms of its gross national product—28 billion dollars. In the process, the country underwent a period of development with inflation, from the end of the Great Depression until 1956, followed by a period of growth with relative price stability that has persisted to the present.

Many of the obstacles to economic development have been overcome through the creation of an extensive base of social overhead capital and the establishment of institutional mechanisms which make for greater mobility of the factors of production, an essential condition for steady growth. Nevertheless, these factors are insufficient to guarantee the continuation of the process of growth. As economic progress takes place, it requires new social overhead capital projects, institutional changes, and the creation of instruments and incentives within economic policy which will enlarge and improve the orientation of investment, as well as the development of skills and the allocation of labor.

The first section presents an introductory description of the outstanding features of the economic behavior and of the nature of economic policy in Mexico since World War II, focusing on topics such as the agricultural sector, the industrial sector, the distribution of income, the financial system, the public finances, and the balance of payments. The purpose of this material is to give the reader a general background with which to appraise the subsequent sections, where opinions on domestic economic policy sustained by Mexican economists, and comments on the same topics by other experts whose views have been influential during the same period, are presented. The second and third sections summarize the views on certain aspects of agricultural and industrial supply and their elasticity. The fourth section examines aspects of allocation, the distribution of income, and consumption and marketing. The fifth and sixth sections describe monetary and fiscal policy respectively. The seventh section summarizes points of view which have been expressed in relation to the foreign sector and the eighth relates to economic development. The ninth section briefly explains the economic planning approaches that have been suggested in Mexico. In the tenth and final section, the author expresses his views on the opinions presented, and on some areas not covered in

---

* Head, Department of Economic Studies, Bank of Mexico, and Professor of Economics, El Colegio de Mexico.

THE AMERICAN ECONOMIC REVIEW

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors

BARBARA R. BERGMANN
JAGDISH N. BHAGWATI
PHILLIP CAGAN
GREGORY C. CHOW
CARL F. CHRIST
C. E. FERGUSON
HARRY G. JOHNSON
DANIEL MCFADDEN
ALVIN L. MARTY
HERBERT MOHRING
MARC NERLOVE
G. WARREN NUTTER
EDMUND S. PHELPS
VERNON L. SMITH

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

• Copyright American Economic Association 1971.

September 1971

VOLUME LXI, NUMBER 4

Articles

The 1971 Report of the President's Council of Economic Advisers: Inflation and Recession *Martin J. Bailey* 517

The 1971 Report of the President's Council of Economic Advisers: Inflation and Recession *Robert Eisner* 522

The 1971 Report of the President's Council of Economic Advisers: Priorities and Efficiency *Abba P. Lerner* 527

The 1971 Report of the President's Council of Economic Advisers: Micro-Economic Aspects of Public Policy *Jerome L. Stein* 531

The Duopoly Game: Output Variations *D. K. Osborne* 538

The Private and Social Value of Information and the Reward to Inventive Activity *Jack Hirshleifer* 561

Entrepreneurial Income: How Does It Measure Up? *Laurits R. Christensen* 575

Substitution, Complementarity, and the Residual Variation Around Dynamic Demand Equations *Louis Phlips* 586

Debt in the Consumer Portfolio: Evidence from a Panel Study *William C. Dunkelberg and Frank P. Stafford* 598

Estimation of Structure-Profit Relationship with Application to the Food Processing Sector *Blake Imel and Peter Helmberger* 614

The Pareto Optimal Approach to Income Redistribution: A Fiscal Application *George M. von Furstenberg and Dennis C. Mueller* 628

Stability in Periodic Markets *Norman P. Obst* 638

Officer Supply—The Impact of Pay, the Draft, and the Vietnam War *Stuart H. Altman and Robert J. Barro* 649

Communications

The Market for a Semipublic Good: The Case of the *American Economic Review* *Yoram Barzel* 665

The Neo-Factor Proportions Theory of International Trade: An Empirical Investigation *Thomas C. Lowinger* 675

Consumer's Surplus and Producer's Goods *Richard Schmalensee* 682

Import Controls on Foreign Oil: Tariff or Quota? *George A. Hay* 688

Profits of Defense Contractors *George J. Stigler and Claire Friedland* 692

On the Empirical Relevancy of the Ranis-Fei Model of Economic Development:
- Comment *Hakchung J. Choo* 695
- Reply *John C. H. Fei and Gustav Ranis* 704

Incentive and Disincentive Experimentation for Income Maintenance Policy Purposes:
- Note *Edgar K. Browning* 709

Pecuniary Externalities: A Game Theoretic Analysis *Martin Shubik* 713

On the Economics of Mass Demonstrations: A Case Study of the November 1969 March on Washington *Charles J. Cicchetti, A. Myrick Freeman III, Robert H. Haveman, and Jack L. Knetsch* 719

The Graduated Fair Return:
- Comment *C. Robert Wichers* 725
- A Further Comment *Alvin K. Klevorick* 727

Mass Balance, General Equilibrium, and Environmental Externalities *Roger G. Noll and John Trijonis* 730

Externalities as Commodities *Robert A. Meyer, Jr.* 736

Stochastic Reserve Losses and Expansion of Bank Credit: Note *J. Phillip Cooper* 741

Discrimination and Income Differentials:
- Comment *Orley Ashenfelter and Michael K. Taussig* 746
- Reply *James Gwartney* 751

Notes 756

THE AMERICAN ECONOMIC REVIEW

December 1971

VOLUME LXI, NUMBER 5

GEORGE H. BORTS
Managing Editor

WILMA ST. JOHN
Assistant Editor

Board of Editors
BARBARA R. BERGMANN
JAGDISH N. BHAGWATI
PHILLIP CAGAN
GREGORY C. CHOW
CARL F. CHRIST
C. E. FERGUSON
HARRY G. JOHNSON
DANIEL MCFADDEN
ALVIN L. MARTY
HERBERT MOHRING
MARC NERLOVE
G. WARREN NUTTER
EDMUND S. PHELPS
VERNON L. SMITH

Articles

Money and Barter in General Equilibrium with Transactions Costs *Jürg Niehans* 773

The Uses of Money: Money in the Theory of an Exchange Economy *Karl Brunner and Allan H. Meltzer* 784

The Supply of Rental Housing *Frank de Leeuw and Nkanta F. Ekanem* 806

The Composition of International Reserve Holdings: A Problem of Choice Involving Risk *John H. Makin* 818

Investment in Graduate Education *Yoram Weiss* 833

Hospital Cost Inflation: A Study of Nonprofit Price Dynamics *Martin S. Feldstein* 853

Congestion Tolls Under Monopoly *Noel M. Edelson* 873

External Diseconomies in Competitive Supply *Charles J. Goetz and James M. Buchanan* 883

Factor Substitution and Effective Protection Reconsidered *Mordechai E. Kreinin, James B. Ramsey, and Jan Kmenta* 891

Deposit Expansion and Federal Reserve-Banking System Interaction: A Micro Unit Simulation *Willard T. Carleton and William R. Bryan* 901

• Manuscripts and editorial correspondence relating to the regular quarterly issue of this REVIEW should be addressed to George H. Borts, Managing Editor of THE AMERICAN ECONOMIC REVIEW, Brown University, Providence, R.I. 02912. Manuscripts should be submitted in duplicate and in acceptable form. *Style Instructions* for guidance in preparing manuscripts will be provided upon request to the editor.

• No responsibility for the views expressed by authors in this REVIEW is assumed by the editors or the publishers, The American Economic Association.

• Copyright American Economic Association 1971.

Communications

Optimal Distributed Lag Responses and Expectations *Roger Craine* 916
The Theory of Tariffs, Taxes, and Subsidies: Some Aspects of the Second Best *Trent J. Bertrand and Jaroslav Vanek* 925
The End of the North-South Wage Differential *Philip R. P. Coelho and Moheb A. Ghali* 932
Subsidies, Economic Lives, and Complete Resource Misallocation *P. Taubman and R. Rasche* 938
Factor Market Imperfections, the Terms of Trade, and Welfare *Raveendra N. Batra and Prasanta K. Pattanaik* 946
Population Density in Patterns of Trade and Development *Donald B. Keesing and Donald R. Sherk* 956
The Elasticity of Expansion and Scale Adjustments of the Firm *Dale B. Truett* 962
Monopsony in the Market for Public School Teachers *John H. Landon and Robert N. Baird* 966
Shortsightedness and Nonmarginal Pareto Optimal Redistribution *A. Mitchell Polinsky* 972
A Set of Explicit Homothetic Production Functions *Benjamin Wolkowitz* 980

Notes 984

Titles of Doctoral Dissertations 990

CALCUTTA UNIVERSITY CENTRAL LIBRARY

SIR W. ARTHUR LEWIS

DISTINGUISHED FELLOW

1970

We honor Sir Arthur Lewis, distinguished scholar and statesman, whose unique analytical skills and ability to understand and persuade have been enlisted time and again in tackling the stubborn problems of poverty. His work on unlimited supplies of labor has been the key to rendering the field of development economics respectable by suggesting the possibility of a new paradigm. His contributions to the planning literature and to actual decision making in the United Kingdom, Ghana, and elsewhere; to higher education in the West Indies; to the counsels of the Pearson Commission; and most recently to development banking in his native Caribbean, all bear testimony to his fundamental concern with the poor and down-trodden. We pay homage to the clarity of his pen, the wit, sometimes caustic, of his lectures, his courage, his insight, and, most of all, his humanity.

W Arthur Lewis

CALC
UNIVERSITY.
CENTRAL
LIBRARY.

THE AMERICAN ECONOMIC REVIEW

VOLUME LXI

BOARD OF EDITORS

Barbara R. Bergmann
Jagdish N. Bhagwati
Phillip Cagan
Gregory C. Chow
Carl F. Christ
C. E. Ferguson
Harry G. Johnson
Daniel McFadden
Alvin L. Marty
Herbert Mohring
Marc Nerlove
G. Warren Nutter
Edmund S. Phelps
Vernon L. Smith

MANAGING EDITOR
GEORGE H. BORTS

THE AMERICAN ECONOMIC ASSOCIATION
Executive Office: Nashville, Tennessee
Editorial Office: Brown University, Providence, Rhode Island

Copyright 1971

AMERICAN ECONOMIC ASSOCIATION

CONTENTS OF ARTICLES AND COMMUNICATIONS

W. Leontief: Theoretical Assumptions and Nonobserved Facts 1

P. A. Diamond and J. A. Mirrlees: Optimal Taxation and Public Production: I—Production Efficiency 8

R. G. Gregory: United States Imports and Internal Pressure of Demand: 1948–68 28

P. K. Bardhan and T. N. Srinivasan: Crop-sharing Tenancy in Agriculture: A Theoretical and Empirical Analysis 48

A. Sandmo: On the Theory of the Competitive Firm Under Price Uncertainty 65

J. C. Leith: The Effect of Tariffs on Production, Consumption, and Trade: a Revised Analysis 74

R. J. Barro and H. I. Grossman: A General Disequilibrium Model of Income and Employment 82

L. J. Lau and P. A. Yotopoulos: A Test for Relative Efficiency and Application to Indian Agriculture 94

J. A. Brittain: The Incidence of Social Security Payroll Taxes 110

R. E. Baldwin: Determinants of the Commodity Structure of U.S. Trade 126

F. Gehrels: Optimal Restrictions on Foreign Trade and Investment 147

D. R. Kamerschen: Large Industrial Corporations and Asset Shares: Comment 160

S. E. Boyle: Comment 163

D. Mermelstein: Reply 168

E. Sheshinski: Welfare Aspects of a Regulatory Constraint: Note 175

M. L. Ladenson: Pitfalls in Financial Model Building: Some Extensions 179

D. G. Davies: Clothing Exemptions and Sales Tax Regressivity: Note 187

H. Theil: An Economic Theory of the Second Moments of Disturbances of Behavioral Equations 190

T. E. Borcherding: A Neglected Social Cost of a Voluntary Military 195

A. O. Converse: On the Extension of Input-Output Analysis to Account for Environmental Externalities 197

M. Krauss and D. M. Winch: Mishan on the Gains from Trade: Comment 199

E. J. Mishan: Reply 202

R. Rosenberg: Profit Constrained Revenue Maximization: Note 208

I. Pressman and A. Carol: Behavior of the Firm Under Regulatory Constraint: Note 210

J. C. Hause: Spectral Analysis and the Detection of Lead-Lag Relations 213

G. Tullock: Subsidized Housing in a Competitive Market: Comment 218

E. O. Olsen: Reply 220

R. Roll: Expectations and the Demand for Bonds: Comment 225

A. Buse: Comment 229

R. A. Kessel: Comment 231

J. H. Wood: Reply 232

A. R. Shepherd: Output of the Restrained Firm: Comment 237

M. Z. Kafoglis: Reply 240

D. B. Stewart: Production Inteterminacy with Three Goods and Two Factors: A Comment on the Pattern of Trade 241

J. R. Melvin: Reply 245

P. A. Diamond and J. A. Mirrlees: Optimal Taxation and Public Production: II—Tax Rules 261

R. L. Miller: A Short-Term Econometric Model of Textile Industries 279

R. Gronau: Information and Frictional Unemployment 290

S. A. Hoenack: The Efficient Allocation of Subsidies to College Students 302

D. Nichols, E. Smolensky, and T. N. Tideman: Discrimination by Waiting Time in Merit Goods 312

R. J. Zeckhauser: Optimal Mechanisms for Income Transfer 324

E. L. Feige and M. Parkin: The Optimal Quantity of Money, Bonds, Commodity Inventories, and Capital 335

E. T. Haefele: A Utility Theory of Representative Government 350

E. Nadel: International Trade and Capital Mobility 368

P. A. Neher: Peasants, Procreation, and Pensions 380

M. Manove: A Model of Soviet-Type Economic Planning 390

T. R. Saving: Transactions Costs and the Demand for Money 407

W. E. Weber: Interest Rates and the Short-Run Consumption Function 421

D. J. Smyth: Unemployment and Inflation: A Cross-Country Analysis of the Phillips Curve 426

R. D. Portes: Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry: An Alternative Approach 430

D. B. Fields and W. T. Stanbury: Income Taxes and Incentives to Work: Some Additional Empirical Evidence.............. 435

B. Hansen: Fiscal and Monetary Policy Reconsidered: Comment.................. 444

J. H. Hotson: Comment.................. 448

B. Henneberry and J. G. Witte: Comment... 452

K. M. Carlson: Comment................. 454

R. Eisner: Reply........................ 458

D. T. Hulett: More on an Empirical Definition of Money: Note....................... 462

K. L. Gupta: Dependency Rates and Savings Rates: Comment....................... 469

N. A. Adams: Comment.................. 472

N. H. Leff: Reply........................ 476

D. Levhari: A Property of a Closed Linear Model of Production: Note.............. 481

R. F. Fullenbaum, E. W. Carlson, and F. W. Bell: Economics of Production from Natural Resources: Comment.................... 483

V. L. Smith: Reply...................... 488

J. A. Hanson: Employment and Rural Wages in Egypt: A Reinterpretation............ 492

B. Hansen: Reply....................... 500

M. J. Bailey: The 1971 Report of the President's Council of Economic Advisers: Inflation and Recession................... 517

R. Eisner: The 1971 Report of the President's Council of Economic Advisers: Inflation and Recession............................. 522

A. P. Lerner: The 1971 Report of the President's Council of Economic Advisers: Priorities and Efficiency................ 527

J. L. Stein: The 1971 Report of the President's Council of Economic Advisers: Micro-Economic Aspects of Public Policy........... 531

D. K. Osborne: The Duopoly Game: Output Variations............................ 538

J. Hirshleifer: The Private and Social Value of Information and the Reward to Inventive Activity............................ 561

L. R. Christensen: Entrepreneurial Income: How Does It Measure Up?.............. 575

L. Phlips: Substitution, Complementarity, and the Residual Variation Around Dynamic Demand Equations................ 586

W. C. Dunkelberg and F. P. Stafford: Debt in the Consumer Portfolio: Evidence from a Panel Study.......................... 598

B. Imel and P. Helmberger: Estimation of Structure-Profit Relationship with Application to the Food Processing Sector........ 614

G. M. von Furstenberg and D. C. Mueller: The Pareto Optimal Approach to Income Redistribution: A Fiscal Application...... 628

N. P. Obst: Stability in Periodic Markets..... 638

S. H. Altman and R. J. Barro: Officer Supply—The Impact of Pay, the Draft, and the Vietnam War......................... 649

Y. Barzel: The Market for a Semipublic Good: The Case of the *American Economic Review*................................ 665

T. C. Lowinger: The Neo-Factor Proportions Theory of International Trade: An Empirical Investigation....................... 675

R. Schmalensee: Consumer's Surplus and Producer's Goods...................... 682

G. A. Hay: Import Controls on Foreign Oil: Tariff or Quota?....................... 688

G. J. Stigler and C. Friedland: Profits of Defense Contractors...................... 692

H. J. Choo: On the Empirical Relevancy of the Ranis-Fei Model of Economic Development: Comment............................ 695

J. C. H. Fei and G. Ranis: Reply........... 704

E. K. Browning: Incentive and Disincentive Experimentation for Income Maintenance Policy Purposes: Note.................. 709

M. Shubik: Pecuniary Externalities: A Game Theoretic Analysis..................... 713

C. J. Cicchetti, A. M. Freeman III, R. H. Haveman, and J. L. Knetsch: On the Economics of Mass Demonstrations: A Case Study of the November 1969 March on Washington.......................... 719

C. R. Wichers: The Graduated Fair Return: Comment............................ 725

A. K. Klevorick: A Further Comment........ 727

R. G. Noll and J. Trijonis: Mass Balance, General Equilibrium, and Environmental Externalities............................ 730

R. A. Meyer, Jr.: Externalities as Commodities................................ 736

J. P. Cooper: Stochastic Reserve Losses and Expansion of Bank Credit: Note.......... 741

O. Ashenfelter and M. K. Taussig: Discrimination and Income Differentials: Comment.............................. 746

J. Gwartney: Reply...................... 751

J. Niehans: Money and Barter in General Equilibrium with Transaction Costs....... 773

K. Brunner and A. H. Meltzer: The Uses of Money: Money in the Theory of an Exchange Economy....................... 784

F. de Leeuw and N. F. Ekanem: The Supply of Rental Housing..................... 806

J. H. Makin: The Composition of International Reserve Holdings: A Problem of Choice Involving Risk.................. 818
Y. Weiss: Investment in Graduate Education.................................. 833
M. S. Feldstein: Hospital Cost Inflation: A Study of Nonprofit Price Dynamics....... 853
N. M. Edelson: Congestion Tolls Under Monopoly.............................. 873
C. J. Goetz and J. M. Buchanan: External Diseconomies in Competitive Supply...... 883
M. E. Kreinin, J. B. Ramsey, and J. Kmenta: Factor Substitution and Effective Protection Reconsidered....................... 891
W. T. Carleton and W. R. Bryan: Deposit Expansion and Federal Reserve-Banking System Interaction: A Micro Unit Simulation.................................. 901
R. Craine: Optimal Distributed Lag Responses and Expectations....................... 916
T. J. Bertrand and J. Vanek: The Theory of Tariffs, Taxes, and Subsidies: Some Aspects of the Second Best...................... 925
P. R. P. Coelho and M. A. Ghali: The End of the North-South Wage Differential........ 932
P. Taubman and R. Rasche: Subsidies, Economic Lives, and Complete Resource Misallocation................................ 938
R. N. Batra and P. K. Pattanaik: Factor Market Imperfections, the Terms of Trade, and Welfare............................ 946
D. B. Keesing and D. R. Sherk: Population Density in Patterns of Trade and Development................................. 956
D. B. Truett: The Elasticity of Expansion and Scale Adjustments of the Firm............ 962
J. H. Landon and R. N. Baird: Monopsony in the Market for Public School Teachers..... 966
A. M. Polinsky: Shortsightedness and Nonmarginal Pareto Optimal Redistribution... 972
B. Wolkowitz: A Set of Explicit Homothetic Production Functions................... 980

CONTENTS OF PAPERS AND PROCEEDINGS

Richard T. Ely Lecture

H. G. Johnson: The Keynesian Revolution and the Monetarist Counter-Revolution 1

Current Status of Income Maintenance Experiments

H. W. Watts: The Graduated Work Incentive Experiments: Current Progress 15

M. Kurz and R. G. Spiegelman: The Seattle Experiment: The Combined Effect of Income Maintenance and Manpower Investments 22

T. F. Kelly and L. Singer: The Gary Income Maintenance Experiment: Plans and Progress 30

Discussion by **J. N. Morgan** 39

The State of Economics: The Behavioral and Social Sciences Survey

H. W. Riecken: Introductory Remarks 43

C. L. Schultze: The Reviewers Reviewed 45

J. G. Gurley: The State of Political Economics 53

Discussion by **R. M. Solow R. L. Heilbroner, and H. W. Riecken** 63

Television: Old Theories, Current Facts, and Future Policies

R. E. Park: The Growth of Cable TV and Its Probable Impact on Over-The-Air Broadcasting 69

J. L. Peterman: Concentration of Control and the Price of Television Time 74

H. J. Levin: Program Duplication, Diversity, and Effective Viewer Choices: Some Empirical Findings 81

E. Greenberg and H. J. Barnett: TV Program Diversity—New Evidence and Old Theories 89

Discussion by **J. J. McGowan and M. J. Peck** 94

Conglomerate and Vertical Responses to Market Imperfection

J. Lintner: Expectations, Mergers and Equilibrium in Purely Competitive Securities Markets 101

O. E. Williamson: The Vertical Integration of Production: Market Failure Considerations 112

Discussion by **R. N. McKean and J. F. Weston** 124

Micro Aspects of Macro Performance

E. C. Budd and D. F. Seiders: The Impact of Inflation on the Distribution of Income and Wealth 128

A. H. Packer: The Two-Way Relationship between the Budget and Economic Variables 139

Discussion by **C. E. Metcalf and J. L. Palmer** 150

The Political Economy of Environmental Quality

A. V. Kneese: Environmental Pollution: Economics and Policy 153

Discussion by **K. E. Boulding, J. W. Carlson, R. W. Judy, and M. J. Roberts** 167

Commercial Policy and Less Developed Countries

B. Balassa: Trade Policies in Developing Countries 178

C. Clague: Tariff Preferences and Separable Utility 188

I. Walter: Nontariff Barriers and the Export Performance of Developing Economies 195

Discussion by **H. B. Malmgren, S. Guisinger, and R. E. Baldwin** 206

Evaluation of Economic Regulation of Industry

F. M. Westfield: Methodology of Evaluating Economic Regulation 211

B. L. Nelson: Costs and Benefits of Regulating Communications 218

A. F. Friedlaender: The Social Costs of Regulating Railroads 226

Discussion by **A. E. Kahn, L. F. Darby and J. R. Nelson** 235

Teaching Economics: Experiments and Results

P. Saunders: The Lasting Effects of Elementary Economics Courses: Some Preliminary Results 242

R. G. Lloyd: Teaching Economics to Black Students 249

B. Castro: Videotaped Dialogues in Economics 256

Discussion by **I. Morrissett, M. L. Halvorsen, and C. R. McConnell** 260

Educational Production Relationships

H. Gintis: Education, Technology, and the Characteristics of Worker Productivity 266

E. Hanushek: Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data 280

J. C. Hause: Ability and Schooling as Determinants of Lifetime Earnings or If You're So Smart, Why Aren't You Rich? 289

Discussion by **W. T. Garner, R. Alcaly, and C. Z. Wilson** 299

Round Table Session on the Academic Labor Market

A. M. Cartter: Whither the Market for Academic Economists? 305
L. R. Harmon: The Supply of Economists in the 1970's 311
F. M. Boddy: Recent Behavior of Economists' Salaries 316
D. G. Brown: Improving the Operation of the Academic Labor Market 321
R. P. Strauss: A Younger Economist's Views of the Market 327

Urban Growth and Development

B. Ellickson: Jurisdictional Fragmentation and Residential Choice 334
E. S. Mills and D. M. de Ferranti: Market Choices and Optimum City Size 340
W. Bateman and H. M. Hochman: Social Problems and the Urban Crisis: Can Public Policy Make a Difference? 346
Discussion by **R. Artle, S. P. Dresch, and A. H. Pascal** 354

The Allocation of Social Risk

W. Brainard and F. T. Dolbear: Social Risk and Financial Markets 360
R. Kihlstrom and M. Pauly: The Role of Insurance in the Allocation of Risk 371
M. Spence and R. Zeckhauser: Insurance, Information, and Individual Action 380
Discussion by **M. J. Bailey, H. P. Minsky, and D. D. Hester** 388

Population and Environment in the United States

J. L. Fisher: Impact of Population on Resources and the Environment 392
R. A. Easterlin: Does Human Fertility Adjust to the Environment? 399
G. G. Cain: Issues in the Economics of a Population Policy for the United States 408
Discussion by **T. P. Schultz and P. Demeny** 418

The Theory of Planning

R. D. Portes: Decentralized Planning Procedures and Centrally Planned Economies 422
T. Nagy: The Hungarian Economic Reform, Past and Future 430
Discussion by **E. Ames and V. Balassa** 436

The Economics of Political Decentralization

D. F. Bradford and W. E. Oates: Towards a Predictive Theory of Intergovernmental Grants 440
A. J. Heins: State and Local Response to Fiscal Decentralization 449
O. A. Davis and K. O. Kortanek: Centralization and Decentralization: The Political Economy of Public School Systems 456
Discussion by **C. J. Goetz, T. E. Borcherding, and W. E. Williams** 463

CONTRIBUTORS TO ARTICLES AND COMMUNICATIONS

Adams, N. A. 472
Altman, S. H. 649
Ashenfelter, O. 746
Bailey, M. J. 517
Baird, R. N. 966
Baldwin, R. E. 126
Bardhan, P. K. 48
Barro, R. J. 82, 649
Barzel, Y. 665
Batra, R. N. 946
Bell, F. W. 483
Bertrand, T. J. 925
Borcherding, T. E. 195
Boyle, S. E. 163
Brittain, J. A. 110
Browning, E. K. 709
Brunner, K. 784
Bryan, W. R. 901
Buchanan, J. M. 883
Buse, A. 229
Carleton, W. T. 901
Carlson, E. W. 483
Carlson, K. M. 454
Carol, A. 210
Choo, H. J. 695
Christensen, L. R. 575
Cicchetti, C. J. 719
Coelho, P. R. P. 932
Converse, A. O. 197
Cooper, J. P. 741
Craine, R. 916
Davies, D. G. 187
de Leeuw, F. 806
Diamond, P. A. 8, 261
Dunkelberg, W. C. 598
Edelson, N. M. 873
Eisner, R. 458, 522
Ekanem, N. F. 806
Fei, J. C. H. 704
Feige, E. L. 335
Feldstein, M. S. 853
Fields, D. B. 435
Freeman, A. M. III, 719
Friedland, C. 692
Fullenbaum, R. F. 483
Gehrels, F. 147
Ghali, M. A. 932
Goetz, C. J. 883
Gregory, R. G. 28
Gronau, R. 290
Grossman, H. I. 82
Gupta, K. L. 469
Gwartney, J. 751
Haefele, E. T. 350
Hansen, B. 444, 500
Hanson, J. A. 492
Hause, J. C. 213
Haveman, R. H. 719
Hay, G. A. 688
Helmberger, P. 614
Henneberry, B. 452
Hirshleifer, J. 561
Hoenack, S. A. 302
Hotson, J. H. 448
Hulett, D. T. 462
Imel, B. 614
Kafoglis, M. Z. 240
Kamerschen, D. R. 160
Kessel, R. A. 231
Keesing, D. B. 956
Klevorick, A. K. 727
Kmenta, J. 891
Knetsch, J. L. 719
Krauss, M. 199
Kreinin, M. E. 891
Ladenson, M. L. 179
Landon, J. H. 966
Lau, L. J. 94
Leff, N. H. 476
Leith, J. C. 74
Leontief, W. 1
Lerner, A. P. 527
Levhari, D. 481
Lowinger, T. C. 675
Makin, J. H. 818
Manove, M. 390
Meltzer, A. H. 784
Melvin, J. R. 245
Mermelstein, D. 168
Meyer, R. A., Jr. 736
Miller, R. L. 279
Mirrlees, J. A. 8, 261
Mishan, E. J. 202
Mueller, D. C. 628
Nadel, E. 368
Neher, P. A. 380
Nichols, D. 312
Niehans, J. 773
Noll, R. G. 730
Obst, N. P. 638
Olsen, E. O. 220
Osborne, D. K. 538
Parkin, M. 335
Pattanaik, P. K. 946
Phlips, L. 586
Polinsky, A. M. 972
Portes, R. D. 430
Pressman, I. 210
Ramsey, J. B. 891
Ranis, G. 704
Rasche, R. 938
Roll, R. 225
Rosenberg, R. 208
Sandmo, A. 65
Saving, T. R. 407
Schmalensee, R. 682
Shepherd, A. R. 237
Sherk, D. R. 956

Sheshinski, E. 175
Shubik, M. 713
Smith, V. L. 488
Smolensky, E. 312
Smyth, D. J. 426
Stafford, F. P. 598
Stanbury, W. T. 435
Stein, J. L. 531
Stewart, D. B. 241
Stigler, G. J. 692
Srinivasan, T. N. 48
Taubman, P. 938
Taussig, M. K. 746
Theil, H. 190
Tideman, T. N. 312
Trijonis, J. 730
Truett, D. B. 962
Tullock, G. 218
Vanek, J. 925
von Furstenberg, G. M. 628
Weber, W. E. 421
Weiss, Y. 833
Wichers, C. R. 725
Winch, D. M. 199
Witte, J. G. 452
Wolkowitz, B. 980
Wood, J. H. 232
Yotopoulos, P. A. 94
Zeckhauser, R. J. 324

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

Alcaly, R. 301
Ames, E. 436
Artle, R. 354
Bailey, M. J. 388
Balassa, B. 178, 437
Baldwin, R. E. 209
Barnett, H. J. 89
Bateman, W. 346
Boddy, F. M. 316
Boulding, K. E. 167
Borcherding, T. E. 463
Bradford, D. F. 440
Brainard, W. 360
Brown, D. G. 321
Budd, E. C. 128
Cain, G. G. 408
Carlson, J. W. 169
Cartter, A. M. 305
Castro, B. 256
Clague, C. 188
Darby, L. F. 237
Davis, O. A. 456
de Ferranti, D. M. 340
Demeny, P. 419
Dolbear, F. T. 360
Dresch, S. P. 355
Easterlin, R. A. 399
Ellickson, B. 334
Fisher, J. L. 392
Friedlaender, A. F. 226
Garner, W. T. 299
Gintis, H. 266
Goetz, C. J. 463
Greenberg, E. 89
Guisinger, S. 207
Gurley, J. G. 53
Halvorsen, M. L. 262
Hanushek, E. 280
Harmon, L. R. 311
Hause, J. C. 289
Heilbroner, R. L. 65
Heins, A. J. 449
Hester, D. D. 390
Hochman, H. M. 346
Johnson, H. G. 1
Judy, R. W. 172
Kahn, A. E. 235
Kelly, T. F. 30
Kihlstrom, R. 371
Kneese, A. V. 153
Kortanek, K. O. 456
Kurz, M. 22
Levin, H. J. 81
Lintner, J. 101
Lloyd, R. G. 249
McConnell, C. R. 264
McGowan, J. J. 94
McKean, R. N. 124
Malmgren, H. B. 206
Metcalf, C. E. 150
Mills, E. S. 340
Minsky, H. P. 389
Morgan, J. N. 39
Morrissett, I. 260
Nagy, T. 430
Nelson, B. L. 218
Nelson, J. R. 239
Oates, W. E. 440
Packer, A. H. 139
Palmer, J. L. 150
Park, R. E. 69
Pascal, A. H. 357
Pauly, M. 371
Peck, M. J. 94
Peterman, J. L. 74
Portes, R. D. 422

Riecken, H. W. 43, 67
Roberts, M. J. 173
Saunders, P. 242
Schultz, T. P. 418
Schultze, C. L. 45
Seiders, D. F. 128
Singer, L. 30
Solow, R. M. 63
Spence, M. 380
Spiegelman, R. G. 22
Strauss, R. P. 327
Walter, I. 195
Watts, H. W. 15
Westfield, F. M. 211
Weston, J. F. 125
Williams, W. E. 465
Williamson, O. E. 112
Wilson, C. Z. 302
Zeckhauser, R. 380

Optimal Taxation and Public Production II: Tax Rules

By Peter A. Diamond and James A. Mirrlees*

In Part I of this paper which appeared in the March 1971 issue of this *Review*, we set out the problem of using taxation and government production to maximize a social welfare function. We derived the first-order conditions, and considered the argument for efficiency in aggregate production. Here in Part II we consider the structure of optimal taxes in more detail. Part I contained five sections, and Part II begins at Section VI. In the sixth and seventh sections we consider commodity taxation in one- and many-consumer economies. In the eighth section we consider other kinds of taxes; and in the ninth, public consumption. In the tenth section we consider a rigorous treatment of the problem, giving a sufficient condition for the validity of the first-order conditions. To begin, we shall restate the notation and basic problem.

Notation

p	producer prices
q	consumer prices
t	taxes ($t=q-p$)
$x^h(q)$	net demand by consumer h (incomes are assumed to equal zero) $h=1, 2, \dots, H$
$u^h(x^h)$	utility function of consumer h
$v^h(q)$	indirect utility function of consumer h $v^h(q)=u^h(x^h(q))$
$X(q)$	aggregate net demand $X(q)=\sum_h x^h(q)$
$U(x^1, \dots, x^H)$	social welfare function
$V(q)$	indirect social welfare function $V(q)=U(x^1(q), \dots, x^H(q))$
$W(u^1, \dots, u^H)$	special case of an individualistic social welfare function, assumed for some of the analysis below.

With this notation before us again, we can restate the welfare maximization problem as that of selecting q to

$$\begin{aligned} &\text{Maximize } V(q) \\ &\text{subject to } G(X(q)) \leqq 0 \end{aligned} \tag{33}$$

where G represents the aggregate production constraint. This problem gave rise to the first-order conditions ((19) and (22)) which were equivalently stated as

$$\begin{aligned} \frac{\partial V}{\partial q_k} &= \lambda \sum p_i \frac{\partial X_i}{\partial q_k} \\ &= -\lambda \frac{\partial}{\partial t_k}\left(\sum t_i X_i\right) \\ &\qquad (k = 1, 2, \dots, n) \end{aligned} \tag{34}$$

Equations (34) were derived only for $k=2, \dots, n$. But we can see that they hold also for $k=1$; for, on multiplying by q_k and adding, we have

$$\sum_{k=1}^{n} \left[\frac{\partial V}{\partial q_k} - \lambda \sum_i p_i \frac{\partial X_i}{\partial q_k}\right] q_k = 0$$

by the homogeneity of degree 0 of V and the X_i. Equation (34) states that the impact of a price rise on social welfare is proportional to the cost of meeting the change in demand induced by the price rise. Alternatively the impact of a tax increase on social welfare is proportional to the induced change in tax revenue (all calculated at fixed producer prices).

---

* Massachusetts Institute of Technology and Nuffield College, respectively. The remainder of the matching footnote in Part I is appropriate here too.

VI. Optimal Tax Structure—One-Consumer Economy

For one consumer and an individualistic welfare function (so that V coincides with v, the indirect utility function of the only consumer in the economy), we can express directly the derivative of social welfare with respect to q_k ($v_k = -\alpha x_k$ where α is the marginal utility of income—see equation (5) of Part I). For this case we can then explore the structure of taxation in more detail. The formulation of the first-order conditions using compensated demand derivatives is due to Paul Samuelson (1951). We begin by stating the familiar Slutsky equation:

$$\frac{\partial x_i}{\partial q_k} = s_{ik} - x_k \frac{\partial x_i}{\partial I} \tag{35}$$

where s_{ik} is the derivative of the compensated demand curve for i with respect to q_k, and $\partial x_i/\partial I$ is the derivative of the uncompensated demand with respect to income (evaluated at $I=0$ in our case). We shall make use of the well-known result that $s_{ik} = s_{ki}$.

Substituting into the first-order conditions (34) we have:

$$\begin{aligned} -\alpha x_k &= -\lambda \frac{\partial}{\partial t_k}\left(\sum t_i x_i\right) \\ &= -\lambda\left(x_k + \sum t_i \frac{\partial x_i}{\partial t_k}\right) \\ &= -\lambda x_k - \lambda \sum t_i s_{ik} \\ &\qquad + \lambda x_k \sum t_i \frac{\partial x_i}{\partial I} \\ &\qquad\qquad k = 1, 2, \ldots, n \end{aligned} \tag{36}$$

Rearranging terms, we can write this in the form:

$$\frac{\sum_i t_i s_{ik}}{x_k} = \frac{\alpha + \lambda - \lambda \sum t_i \frac{\partial x_i}{\partial I}}{\lambda} \tag{37}$$

The point to be noticed is that the right-hand side of this equation is independent of k. Call it $-\theta$. Finally, using the symmetry of the Slutsky matrix, we write the first-order conditions as:

$$\frac{\sum_i s_{ki} t_i}{x_k} = -\theta \tag{38}$$

Multiplying by $t_k x_k$ and summing, we obtain

$$\theta \sum_k t_k x_k = -\sum_{k,i} t_k s_{ki} t_i \geq 0, \tag{39}$$

by the negative semi-definiteness of the Slutsky matrix. Thus θ has the same sign as net government revenue.

The left-hand side of (38) is the percentage change in the demand for good k that would result from the tax change if producer prices were constant, the consumer were compensated so as to stay on the same indifference curve, and the derivatives of the compensated demand curves were constant at the same level as at the optimum point:

$$\begin{aligned} \Delta x_k &= \sum_i \int_0^{t_i} \frac{\partial x_k}{\partial t_i} dt_i = \sum_i \int_0^{t_i} s_{ki} dt_i \\ &= \sum_i s_{ki} \int_0^{t_i} dt_i = \sum_i s_{ki} t_i \end{aligned} \tag{40}$$

In fact, it is not possible for all these derivatives to be constant. But if the optimal taxes are small, it is approximately true that the optimal tax structure implies an equal percentage change in compensated demand at constant producer prices.

We can also calculate the actual changes in demand arising from the tax structure (assuming price derivatives of demand and production prices are constant) by resubstituting from the Slutsky equation (35). Then, upon substitution, we have:

$$\sum_i \frac{\partial x_k}{\partial q_i} t_i + \frac{\partial x_k}{\partial I} \sum t_i x_i = -\theta x_k;$$

or

$$\frac{\sum_i \frac{\partial x_k}{\partial q_i} t_i}{x_k} = -\theta - x_k^{-1} \frac{\partial x_k}{\partial I} \sum t_i x_i \tag{41}$$

The actual changes in demand (again assuming constant derivatives) induced by the tax structure differ from proportionality with a larger than average percentage fall in demand for goods with a large income derivative.

Three-Good Economy

In the case of a three-good economy, we can obtain an expression for the relative ad valorem tax rates of the two taxed goods. This argument is similar to that of W. J. Corlett and D. C. Hague, who discussed the direction of movement away from proportional taxation that would increase utility. In the three-good case, with good one untaxed, the first-order conditions (38) become

$$\begin{aligned} s_{22}t_2 + s_{23}t_3 &= -\theta x_2 \\ s_{32}t_2 + s_{33}t_3 &= -\theta x_3 \end{aligned} \tag{42}$$

Solving these equations we have

$$t_2 = \theta \frac{s_{23}x_3 - s_{33}x_2}{s_{22}s_{33} - s_{23}^2}, \quad t_3 = \theta \frac{s_{32}x_2 - s_{22}x_3}{s_{22}s_{33} - s_{23}^2} \tag{43}$$

Notice that the denominator here is positive, by the properties of the Slutsky matrix. We convert these into elasticity expressions, defining the elasticity of compensated demand by

$$\sigma_{ij} = \frac{q_j s_{ij}}{x_i} \tag{44}$$

Equation (43) can then be written

$$\frac{t_2}{q_2} = \theta'(\sigma_{23} - \sigma_{33}), \quad \frac{t_3}{q_3} = \theta'(\sigma_{32} - \sigma_{22}), \tag{45}$$

where

$$\theta' = \frac{\theta x_2 x_3}{q_2 q_3 (s_{22}s_{33} - s_{23}^2)}$$

We now substitute for σ_{23} and σ_{33}, using the adding-up properties of compensated elasticities,

$$\begin{aligned} \sigma_{23} &= -\sigma_{22} - \sigma_{21}, \\ \sigma_{32} &= -\sigma_{33} - \sigma_{31} \end{aligned} \tag{46}$$

This gives us

$$\begin{aligned} \frac{t_2}{q_2} &= \theta'(\sigma_{21} + \sigma_{22} + \sigma_{33}), \\ \frac{t_3}{q_3} &= \theta'(\sigma_{31} + \sigma_{22} + \sigma_{33}) \end{aligned} \tag{47}$$

The interesting case to consider is where labor ($x_1 < 0$) is the untaxed good, while goods 2 and 3 are consumer goods ($x_2 > 0$, $x_3 > 0$). Then θ' has the same sign as net government revenue. For definiteness, suppose that government revenue is positive so that $\theta' > 0$. Equation (47) shows that

$$\frac{t_2}{q_2} \gtreqless \frac{t_3}{q_3} \text{ according as } \sigma_{21} \lesseqgtr \sigma_{31} \tag{48}$$

The tax rate is proportionally greater for the good with the smaller cross-elasticity of compensated demand with the price of labor. (It is possible that one commodity is subsidized, but it has to be the one with the greater cross-elasticity.)

Examples

The implications of the above model are very diverse, depending upon the nature of the demand functions. A simple example will show how the theory can be used. If we define ordinary demand elasticities by the usual formula

$$e_{ik} = q_k x_i^{-1} \frac{\partial x_i}{\partial q_k}, \tag{49}$$

we can rewrite the optimal taxation formula in the form

$$v_k = q_k^{-1} \lambda \sum p_i x_i e_{ik} \tag{50}$$

When the welfare function is individualistic, equation (5) applies, so that equation (50) may be written as

$$-\alpha q_k x_k = \lambda \sum p_i x_i \epsilon_{ik} \tag{51}$$

or

$$q_k p_k^{-1} = -\frac{\lambda}{\alpha} \sum_i \frac{p_i x_i}{p_k x_k} \epsilon_{ik}$$

If we have a good whose price does not affect other demands (implying a unitary own price elasticity), equation (51) simplifies to yield the optimal tax of that good:

$$\begin{aligned} &\text{If } \epsilon_{ik} = 0 \ (i \neq k) \quad \text{and} \quad \epsilon_{kk} = -1, \\ &\text{then } q_k p_k^{-1} = \lambda \alpha^{-1} \end{aligned} \tag{52}$$

where $q_k p_k^{-1}$ equals one plus the percentage tax rate. Recalling that α is the marginal utility of income while λ reflects the change in welfare from allowing a government deficit financed from some outside source, their ratio gives a marginal cost (in terms of the numeraire good) of raising revenue. Thus the optimal tax rate on such a good gives the cost to society of raising the marginal dollar of tax.

An example of a utility function exhibiting such demand curves is the Cobb-Douglas, where only labor is supplied. As an example consider:

$$u(x) = b_1 \log(x_1 + \omega_1) + \sum_{i=2}^{n} b_i \log x_i \tag{53}$$

If we choose labor as the untaxed numeraire, all other goods satisfy (52) and we see that the optimal tax structure is a proportional tax structure.

It is easy to exhibit examples where the optimal tax structure is not proportional. Consider the example:

$$\begin{gathered} u(x) = \sum b_i \log(x_i + \omega_i), \\ \sum b_i = 1, \ \omega_i \neq 0 \end{gathered} \tag{54}$$

The demands arising from these preferences are:

$$x_i = q_i^{-1} b_i \sum q_j \omega_j - \omega_i \tag{55}$$

Therefore the demand elasticities are:

$$\begin{aligned} \epsilon_{ik} &= b_i \omega_k x_i^{-1} \frac{q_k}{q_i} \qquad (k \neq i) \\ \epsilon_{kk} &= -b_k x_k^{-1} \sum_{j \neq k} \omega_j \frac{q_j}{q_k} \end{aligned} \tag{56}$$

Substituting in the formula for the optimal taxes,

$$\begin{aligned} -\alpha q_k x_k &= \\ &\lambda \left[\sum_{j \neq k} b_j \frac{p_j}{q_j} \omega_k q_k - b_k \frac{p_k}{q_k} \sum_{j \neq k} \omega_j q_j \right] \\ &= \lambda \sum_j \left[b_j \omega_k \frac{p_j q_k}{q_j} - b_k \omega_j \frac{p_k q_j}{q_k} \right] \end{aligned} \tag{57}$$

Since the assumption $\sum b_j = 1$ allows us to write the demand functions (55) in the form:

$$q_k x_k = \sum_j \left[b_k \omega_j q_j - b_j \omega_k q_k \right], \tag{58}$$

we can deduce from (57) and (58) that

$$\begin{aligned} \sum_j \Bigg[& b_j \omega_k q_k \left(\frac{p_j}{q_j} - \frac{\alpha}{\lambda} \right) \\ & - b_k \omega_j q_j \left(\frac{p_k}{q_k} - \frac{\alpha}{\lambda} \right) \Bigg] = 0 \end{aligned} \tag{59}$$

These equations allow us to calculate p for any given q, and in that way give the optimal taxation rules. In general, taxes will not be proportional. As one example of this, consider the following three-good case.

Sample Calculation

Let us combine the above two examples by considering a three-good economy (one-consumer good and two types of labor) with preferences as in (54). This example will be used to show that limited tax possibilities (represented by the same proportional tax on goods 2 and 3) introduces the desirability of aggregate production inefficiency.

Example e. Assume that preferences satisfy

$$u = \log x_1 + \log (x_2 + 1) + \log (x_3 + 2) \tag{60a}$$
$$x_1 > 0, \quad x_2 > 1, \quad x_3 > -2;$$

while private production possibilities are

$$y_1 + y_2 + y_3 \leqq 0, \tag{60b}$$
$$y_1 \geqq 0, \quad y_2 \leqq 0, \quad y_3 \leqq 0;$$

and the government constraint is

$$1.02z_1 + z_2 \leqq 0 \tag{60c}$$
$$z_1 \geqq 0, \quad z_2 \leqq 0, \quad z_3 \leqq -0.1$$

Thus the government needs good 3 for public use and can produce good 1 from good 2, but only less efficiently than the private sector can.

Since we know that production efficiency is desired, we have

$$q_1 = p_1 = p_2 = 1, \qquad z_1 = z_2 = 0$$

From the first-order conditions (59) and market clearance given the demands (58), we obtain two equations to determine q_2 and q_3:

$$q_2(q_3^{-1} - 1) = 2q_3(q_2^{-1} - 1)$$
$$(q_2 + 2q_3)(q_2^{-1} + q_3^{-1} + 1) = 8.7$$

These have a unique positive solution

$$q_2 = 0.94494, \quad q_3 = 0.90008$$

which give

$$x_1 = 0.9150, \quad x_2 = -0.0316, \quad x_3 = -0.9834$$
$$u = -0.1045$$

If we now require the same tax rate on goods 2 and 3 and at the same time impose production efficiency, then $q_2 = q_3 = q$, and the tax rate is determined by the market clearance equation. We obtain

$$3q + 6 = 8.7; \qquad \text{i.e., } q = 0.9$$

Then demands are

$$x_1 = 0.9, \quad x_2 = 0, \quad x_3 = -1$$

and

$$u = -0.1054$$

Notice that the economy is still on the production frontier even though both input prices are lower in this case. If we introduce inefficiency with $p_2 > 1$, so that $y_2 = 0$ and $x_2 = z_2$, we can increase utility. Market clearance now requires

$$(q_2 + 2q_3)((1.02)^{-1}q_2^{-1} + q_3^{-1} + 1) = 8.7$$

At prices $q_2 = .92$, $q_3 = .90008$ for example, we have, $x_1 = 0.9067$, $x_2 = -0.0144$, $x_3 = -0.9926$, and $u = -0.1051$.

VII. Optimal Tax Structure—Many-Consumer Economy

As we noted in Section III of Part I, the equations for optimal taxation with a single consumer which do not reflect the particular form of V are also valid for many consumers. To pursue the analysis further, we must find an expression for V_k, the derivative of social welfare with respect to the kth consumer price.

With an individualistic welfare function, we have

$$V(q) = W(v^1(q), v^2(q), \ldots, v^H(q)) \tag{61}$$

Differentiating with respect to q_k, we obtain

$$V_k = \sum_h \frac{\partial W}{\partial u^h} v_k^h = -\sum_h \frac{\partial W}{\partial u^h} \alpha^h x_k^h \tag{62}$$

The term α^h is the marginal utility of income of consumer h. Therefore

$$\beta^h = \frac{\partial W}{\partial u^h} \alpha^h \tag{63}$$

is the increase in social welfare from a unit increase in the income of consumer h. We have

$$-V_k = \sum_h \beta^h x_k^h, \tag{64}$$

or the derivative of welfare with respect to a price equals the "welfare-weighted" net consumer demand for commodity k. The necessary condition for optimal taxation makes V_k proportional to the marginal contribution to tax revenue from raising the tax on good k.

$$\sum_h \beta^h x_k^h = \lambda \frac{\partial T}{\partial t_k}, \tag{65}$$

where $T = \sum t_i X_i$ is total tax revenue, and the derivative is evaluated at constant producer prices (i.e., on the basis of consumer excess demand functions alone). We also have the alternative formula

$$\sum_h \beta^h x_k^h = -\lambda \sum_i p_i \frac{\partial X_i}{\partial q_k} \tag{66}$$

Example f. Before turning to interpretations of the optimal tax formulae like those above, let us consider an example.

We will assume that each consumer has a Cobb-Douglas utility function,

$$\begin{aligned} u^h &= b_1^h \log (x_1^h + \omega^h) \\ &\quad + \sum_2^n b_i^h \log x_i^h, \qquad \sum_1^n b_i^h = 1 \end{aligned} \tag{67}$$

Choosing good 1 as numeraire, we saw in Section VI that with a one-consumer economy, taxation would be proportional. This will not, in general, be true in a many-consumer economy where each consumer has this utility function. The individual demand curves arising from this utility function are:

$$\begin{aligned} x_i^h &= q_i^{-1} b_i^h q_1 \omega^h, \qquad i = 2, 3, \ldots, n \\ x_1^h &= -(1 - b_1^h)\omega^h \end{aligned} \tag{68}$$

Notice that $\partial x_i^h / \partial q_k = 0 \ (k \neq i \neq 1)$ and $\partial x_i^h / \partial q_i = -x_i^h / q_i \ (i \neq 1)$.

Assuming an individualistic welfare function, the first-order conditions (66) are in this case

$$\sum_h \beta^h x_k^h = \lambda p_k q_k^{-1} \sum_h x_k^h \tag{69}$$

$$(k = 2, \ldots, n)$$

This implies the following formula:

$$\frac{q_k}{p_k} = \lambda \frac{\sum_h x_k^h}{\sum_h \beta^h x_k^h} = \lambda \frac{\sum_h b_k^h \omega^h}{\sum_h \beta^h b_k^h \omega^h} \tag{70}$$

$$(k = 2, \ldots, n)$$

To complete the determination of the optimal taxes, we must find the relationship between λ, p_1, and q_1. This is obtained from the Walras identity. The value of net consumer demand in producer prices is equal to minus the profit in production. (Alternatively, we could determine λ so that the government budget is balanced.) That is

$$\begin{aligned} &-p_1 \sum_h (1 - b_1^h)\omega^h \\ &\qquad + \sum_{i=2}^n \sum_h p_i q_i^{-1} b_i^h q_1 \omega^h = \gamma, \end{aligned} \tag{71}$$

where γ is the maximized profit of production net of government needs $(= \sum_{i=1}^n p_i z_i)$. Substituting from (70) and rearranging, we obtain

$$\begin{aligned} \frac{q_1}{p_1} &= \lambda \frac{\sum_h (1 - b_1^h)\omega^h + \gamma p_1^{-1}}{\sum_{i=2}^n \sum_h \beta^h b_i^h \omega^h} \\ &= \lambda \frac{\sum (1 - b_1^h)\omega^h + \gamma p_1^{-1}}{\sum \beta^h (1 - b_1^h)\omega^h} \end{aligned} \tag{72}$$

The number γp^{-1} is determined by the technology and the government expenditure decision, and therefore depends on p (unless $\gamma = 0$).

Equations (70) and (72) determine the optimal tax rates. If the social marginal utilities, β^h, are independent of taxation, the optimal tax rates can be read off at once. This is true if W has the special form $\sum_h v^h$; for in that case $\beta^h = 1/\omega^h$. It should be noticed that, although each household's social marginal utility of income is unaffected by taxation, it is desirable to have taxation in general. If households with relatively low social marginal utility of income predominate among the purchasers of a commodity, that commodity should be relatively highly taxed. Although such taxation does nothing to bring social marginal utilities of income closer together, it does increase total welfare.

In general, taxation does affect social marginal utilities of income. The β^h depend on the tax rates, and equations (70) do not, therefore, give explicit formulae for the optimum taxes. In the case $W = -\mu^{-1} \sum_h e^{-\mu v^h}$, $\mu > 0$, so that there is a stronger bias toward equality than in the additive case, it can be verified quite easily that the optimum taxes have to satisfy

$$\begin{aligned} (73) \qquad & \frac{q_k}{p_k} \sum_h b_k^h (\omega^h)^{-\mu} \prod_{i=2}^{n} (b_i^h)^{-\mu b_i^h} q_i^{\mu b_i^h} \\ & = \lambda \sum_h b_k^h \omega^h \qquad (k = 2, 3, \ldots, n) \end{aligned}$$

In this case, marginal utilities of income are brought closer together.[1] It is not immediately obvious from the equations (10) that the q are determined given the p. However, it can be shown that, in the present example, the first-order conditions must have a unique solution.[2] In fact, the relations (70) (along with (72)) would, if followed by government, certainly lead to maximum welfare if production were perfectly competitive, since any state of the economy satisfying these conditions maximizes welfare, and the maximum is unique for the welfare function considered. Unfortunately this convenient property is not general.

From equation (70) we can identify two cases where optimal taxation is proportional. If the social marginal utility of income is the same for everyone ($\beta^h = \beta$, for all h), then equation (70) reduces to $q_k p_k^{-1} = \lambda/\beta$. In this case there is no welfare gain to be achieved by redistributing income, and so no need to tax differently (on average) the expenditures of different individuals. Thus the optimal tax formula has the same form as in the one-consumer case. When the β^h do differ, taxes are greater on commodities purchased more heavily by individuals with a low social marginal utility of income. If, for example, the welfare function treats all individuals symmetrically and if there is diminishing social marginal utility with income, then there is greater taxation on goods purchased more heavily by the rich.

The second case leading to proportional taxation occurs when demand vectors are proportional for all individuals, $x^h = \rho^h x$, and thus $b_k^h = b_k$ for all h. With all individuals demanding goods in the same proportions, it is impossible to redistribute income by commodity taxation implying that the tax structure again assumes the form it has in a one-consumer economy.

Optimal Tax Formulae

The description in Section VI of some possible interpretations of the optimal tax formula carries over to the many-consumer case. Thus, as was true there con-

---

[1] If $\mu < 0$, utilities and marginal utilities are moved further apart.

[2] It is easily verified that $v^h = \delta_h + \sum_i b_i \, log \, (q_1/q_i)$, where the δ_h are constants. Consequently

$$V(q) = -\mu^{-1} \sum_h e^{-\mu\delta_h} \prod{}_i (q_1/q_i)^{-\mu b_i^h}$$

which is a concave function of $(q_1/q_2, q_1/q_2, \ldots, q_1/q_n)$. Also, aggregate demand is

$$X_i(q) = \sum_h b_i^h \omega^h \cdot (q_1/q_i), \; X_1(q) = -\sum_h (1 - \alpha_i^h)\omega^h$$

If the production set is convex, the set of $(q_1/q_2, \ldots, q_1/q_n)$ for which $(X_1, X_2, \ldots, X_n)$ is feasible is also convex. Thus the optimum q is obtained by maximizing a concave function of $(q_1/q_2, \ldots, q_1/q_n)$ over a convex set, and is therefore uniquely defined by the first-order conditions.

sumer price elasticities but not producer price elasticities enter the equations, and at the optimum the social marginal utility of a price change is proportional to the marginal change in tax revenue from raising that tax, calculated at constant producer prices. Analysis of the change in demand can also be carried out, but is naturally more complicated. Assuming an individualistic welfare function, the first-order conditions can be written[3]

$$\sum \beta^h x_k^h = \lambda \sum_h \sum_i t_i \frac{\partial x_i^h}{\partial q_k} + \lambda \sum_h x_k^h \tag{74}$$

From the Slutsky equation, we know that

$$\begin{aligned} \frac{\partial x_i}{\partial q_k} &= s_{ik} - x_k \frac{\partial x_i}{\partial I} = s_{ki} - x_k \frac{\partial x_i}{\partial I} \\ &= \frac{\partial x_k}{\partial q_i} - x_k \frac{\partial x_i}{\partial I} + x_i \frac{\partial x_k}{\partial I} \end{aligned} \tag{75}$$

Substituting from (75) in (74) we can write the optimal tax formula as equation (76). Rearranging terms we can write equation (76) as (77). With constant producer prices, equation (77) gives the change in demand as a result of taxation for a good with constant price-derivatives of the demand function (or for small taxes). Considering two such goods, we see that the percentage decrease in demand is greater for the good the demand for which is concentrated among:

(1) individuals with low social marginal utility of income,
(2) individuals with small decreases in taxes paid with a decrease in income,
(3) individuals for whom the product of the income derivative of demand for good k and taxes paid are large.

VIII. Other Taxes

Thus far we have examined the combined use of public production and commodity taxation as control variables. It is natural to reexamine the analysis when additional tax variables are included in those controlled by the government. In particular, in the next subsection we will briefly consider income taxation; but first, let us examine a general class of taxes such that the consumer budget constraint depends on consumer prices and on tax variables. We shall replace the budget constraint $\sum q_i x_i = 0$ by the more general constraint $\phi(x, q, \zeta) = 0$, where ζ represents a shift parameter to reflect the choice among different systems of additional taxation (for example, the degree of progression in the income tax). Let us note that this formulation continues to assume that all taxes are levied on consumers and that there are no profits in the economy.

The key assumption to permit an extension of the analysis above is an independence of the two constraints on the planner. We need to assume that the choice of tax variables does not affect the production

$$\sum_h \beta^h x_k^h = \lambda \sum_h \sum_i t_i \frac{\partial x_k^h}{\partial q_i} + \lambda \sum_h \sum_i t_i \left(x_i^h \frac{\partial x_k^h}{\partial I} - x_k^h \frac{\partial x_i^h}{\partial I} \right) + \lambda \sum_h x_k^h \tag{76}$$

$$\frac{\sum_h \sum_i t_i \frac{\partial x_i^h}{\partial q_i}}{\sum_h x_k^h} = \frac{1}{\lambda} \frac{\sum_h \beta^h x_k^h}{\sum_h x_k^h} - 1 + \frac{\sum_h \left(\sum_i t_i \frac{\partial x_i^h}{\partial I} \right) x_k^h}{\sum_h x_k^h} - \frac{\sum_h \left(\sum_i t_i x_i^h \right) \frac{\partial x_k^h}{\partial I}}{\sum_h x_k^h} \tag{77}$$

---

[3] We neglect the possibility of a free good when the first-order condition would be an inequality.

possibilities, and further that the choice of a production point does not affect the set of possible demand configurations. In particular, this formulation implies that producer prices do not affect consumer budget constraints. Thus the income tax, to fit this formulation, needs to be levied on the wages that consumers receive, not on the cost of wages to the firm. Similarly it is assumed that there are no sales tax deductions from the income tax base.

We know already that in such a case, optimal production is efficient. We may therefore concentrate upon the case in which all production is controlled by the government, and the production constraint is that $x_1 = g(x_2, x_3, \ldots, x_n)$. We have to choose $q_2, q_3, \ldots, q_n, \zeta$ to

$$\text{(78)} \quad \text{maximize } V(q, \zeta) \text{ subject to } X_1(q, \zeta) = g(X_2(q, \zeta), \ldots, X_n(q, \zeta))$$

As before we introduce a Lagrange multiplier λ. Differentiation with respect to q_k yields the familiar

$$\text{(79)} \qquad V_k = \lambda \sum_i p_i \frac{\partial X_i}{\partial q_k},$$

where the producer price p_i is $\partial g / \partial x_i$ $(i = 2, 3, \ldots, n)$, and $p_1 = 1$. Differentiation with respect to the new tax variable provides the similar equation

$$\text{(80)} \qquad \frac{\partial V}{\partial \zeta} = \lambda \sum_i p_i \frac{\partial X_i}{\partial \zeta}$$

We have an alternative form for (79), namely,

$$\text{(81)} \qquad V_k = -\lambda \frac{\partial T}{\partial t_k}$$

In exactly the same way, we obtain from (80) a formula involving the effect of the new tax on total tax revenue,

$$\text{(82)} \qquad V_\zeta = -\lambda \frac{\partial T}{\partial \zeta}$$

Income Taxation

Nothing that we have said suggests that commodity taxation is superior to income taxation. The analysis has only considered the best use of commodity taxation. It is natural to go on to ask how one employs both commodity taxation and income taxation. The formulation of income taxation raises a problem. If the planners are free to select any income tax structure and if there are a finite number of tax payers, the tax structure can be selected so that the marginal tax rate is zero for each taxpayer at his equilibrium income (although this does not necessarily bring the economy to the full welfare maximum). This eliminates much of our problem, but like lump sum taxation, seems to be beyond the policy tools available in a large economy. The natural formulation of this problem is for a continuum of tax payers, since then no man can have a tax schedule tailor-made for him. (This approach is taken by Mirrlees.) However, we shall here take the alternative route by assuming a limited set of alternatives for the income tax structure.

If only commodity taxation is possible, the tax paid by a household that purchases a vector x^h is

$$\text{(83)} \qquad T^h = \sum_i t_i x_i^h$$

To add income taxation to the tax structure, we can select a subset of commodities, L, e.g., labor services, and tax the value of transactions on this subset, so that

$$I^h = \sum_{i \text{ in } L} q_i x_i^h$$

where I is "taxable income." Then

$$\text{(84)} \qquad T^h = \sum_i t_i x_i^h + \tau(I^h, \zeta),$$

where τ is a fixed continuously differentiable function depending on a parameter ζ, and is the same for all consumers. With a tax on services (x_i negative) we would expect τ to be decreasing in its tax base, with a derivative between zero and minus one. In terms of the notation employed above, we can define the budget constraint $\phi(x^h, q, \zeta)$ by

$$\begin{aligned} &\phi(x^h, q, \zeta) \\ (85)\qquad &= \sum p_i x_i^h + T^h \\ &= \sum q_i x_i^h + \tau\left(\sum_{i \text{ in } L} q_i x_i^h, \zeta\right) \end{aligned}$$

Here we can regard q and ζ as the policy variables. Thus the consumer's budget constraint can be expressed in a form depending on consumer prices and independent of producer prices.

The first-order conditions for optimal income taxation are just the conditions (79) and (80), interpreted for this special case. The social marginal utility of a tax variable change is proportional to the marginal change in tax revenue calculated at constant producer prices. In the case of an individualistic welfare function, we can give more explicit formulae for the welfare derivatives, V_k and V_ζ:

$$(86)\qquad V_k = \sum_h \beta^h x_k^h \left(1 + \delta_k \frac{\partial \tau^h}{\partial I}\right)$$

$$(87)\qquad V_\zeta = \sum_h \beta^h \frac{\partial \tau^h}{\partial \zeta},$$

where $\delta_k = 1$ if k is in L, 0 if k is not in L; and $\tau^h = \tau(I^h, \zeta)$.

These equations are derived from the first-order conditions for maximizing u^h subject to $\phi = 0$, noticing that, for example, the budget constraint implies that

$$\sum_k \frac{\partial \phi}{\partial x_k} \frac{\partial x_k}{\partial \zeta} + \frac{\partial \phi}{\partial \zeta} = 0$$

Combining (82) and (87), we obtain

$$(88)\qquad \sum \beta^h \frac{\partial \tau^h}{\partial \zeta} = \lambda \frac{\partial T}{\partial \zeta}$$

Thus, at the optimum, for any two different kinds of change in the income tax structure, the social-marginal-utility weighted changes in taxation (consumer behavior held constant) are proportional to the changes in total tax revenue (both income and commodity tax revenue, calculated at fixed producer prices, with consumer behavior responding to the price change).

IX. Public Consumption

From the start, we have considered the government production decision as constrained by $G(z) \leqq 0$. The presence of a fixed bundle of public consumption was therefore included in the model (and would show itself by $G(0)$ being positive). This is unsatisfactory and was assumed to keep as uncluttered as possible a naturally complicated problem. We can now consider a choice among vectors of public consumption which affect social welfare directly. (We shall assume that the government controls all production, thus ignoring public expenditures which affect private production rather than consumer utility.) Let us denote by e the vector of public consumption expenditures. (Items of public consumption which are difficult to measure can be described by the inputs into their production.) The presence of public consumption alters our problem in three ways. First, public consumption represents public production (or purchases) which are not supplied to the market. Thus market clearance becomes $X = z - e$.

Second, the presence of public consumption affects private net demand, which must now be written $X(q, e)$. Third, the level of public consumption directly affects the social welfare function (by affecting individual utility in the case of an individualistic welfare function).

We can restate the basic maximization problem as

$$(89)\quad \underset{q,e}{\text{Maximize}}\ V(q, e) \quad \text{subject to } G(X(q, e) + e) \leqq 0$$

The presence of e in the problem will not affect the equations obtained by differentiating a Lagrangian expression with respect to q. Thus the presence of alternative bundles of public consumption does not alter the rules for the optimal tax structure. Nor would we expect it to affect the conditions which imply production efficiency at the optimum. We can therefore replace the inequality in (89) with an equality and differentiate the Lagrangian expression with respect to e_k:

$$(90)\quad \frac{\partial V}{\partial e_k} - \lambda\left[\sum G_i \frac{\partial X_i}{\partial e_k} + G_k\right] = 0$$

Since

$$(91)\quad \sum G_i \frac{\partial X_i}{\partial e_k} = \sum p_i \frac{\partial X_i}{\partial e_k} = \sum (q_i - t_i)\frac{\partial X_i}{\partial e_k} = \frac{\partial}{\partial e_k}\left(\sum q_i X_i - \sum t_i X_i\right) = -\frac{\partial}{\partial e_k}\left(\sum t_i X_i\right),$$

we can write (90) as

$$(92)\quad \frac{\partial V}{\partial e_k} = -\lambda \frac{\partial}{\partial e_k}\left(\sum t_i X_i\right) + \lambda G_k$$

Equations (92) show how the optimal level of public consumption depends on:

(i) the direct contribution of public consumption to welfare (measured by $\partial V/\partial e_k$);

(ii) the effect of public consumption on tax revenue (measured by $\partial \sum t_i X_i/\partial e_k$); and

(iii) the direct cost of public consumption (G_k).

There are three differences between this theory and that of public goods in the presence of lump sum taxation (as developed, for example, by Samuelson (1954)). Because social marginal utilities of income are not equated, the expression $\partial V/\partial e_k$ cannot be reduced to a sum of marginal rates of substitution, but depends on the weights given to the different beneficiaries of public consumption:

$$(93)\quad \frac{\partial V}{\partial e_k} = \sum_h \frac{\partial W}{\partial u^h}\frac{\partial u^h}{\partial e_k}$$

Second, the cost associated with the raising of government revenue implies that the impact of public consumption on revenue is a relevant part of the first-order conditions. Third, for the same reason, the cost of public consumption is measured in terms of the cost to the government of raising revenue to finance the expenditures (in terms of the one-consumer equation, λ may not be equal to α, the marginal utility of income).

The first-order conditions for the provision of public goods can be expressed in another way, showing the relationships between the marginal cost and "willingness to pay." Write r_k^h for the marginal rate of substitution between public good k and income for the hth household. Then $\partial u^h/\partial e_k = \alpha^h r_k^h$, where α^h is the hth household's marginal utility of income. The social marginal utility of the hth household's income, β^h, is $(\partial W/\partial u^h)\alpha^h$. Consequently, from (93)

$$(94)\quad \frac{\partial V}{\partial e_k} = \sum_h \beta^h r_k^h$$

Then, from (92)

$$(95)\quad G_k = \sum_h \left[\frac{\beta^h}{\lambda} r_k^h + \frac{\partial}{\partial e_k}\sum_i t_i x_i^h\right]$$

Thus the marginal cost of producing the public good should be equated to a sum, over all households, of the price which the household is just willing to pay for a marginal increment in the level of provision, weighted by the marginal "social worth" of the household's income, and adjusted for the effect of the level of provision on net tax payments by the household.[4]

In the discussion of public consumption thus far it has been assumed that there were no possible fees associated with the provision of public goods. This would be appropriate for national defense or preventive medicine, but not for goods where licenses can be required from users. The optimal level of license fees will not, in general, be zero. Indeed we may be able to associate with any good more complicated pricing mechanisms than the single fixed price considered above. In particular, there are the familiar examples of two-part tariffs (a license fee for use of a facility plus a per unit charge on the amount of use), and prices depending on quantity of sales. Formally these can be treated in a fashion similar to the income taxes considered above; the set of goods over which the tax is defined is now a consumption good rather than labor. With a two-part tariff, this would imply a tax function which was not continuous at the origin.

Presumably the introduction of more general pricing and taxing schemes gives an opportunity for increasing social welfare, just as the progressive income tax gives such an opportunity. In practice, the ignored costs of tax administration may severely limit the number of complicated pricing schemes which can increase welfare. We would expect the analysis done above to be basically unchanged by the addition of these possibilities, although a two-part tariff will cause aggregate demand to have discontinuities. In practice we would expect these discontinuities to be small relative to aggregate demand, and formally, they could be eliminated by the device of a continuum of consumers.

X. The Optimal Taxation Theorem

In the earlier discussion, we employed calculus techniques to obtain the first-order conditions for the optimal tax structure. However, the valid use of Lagrange multipliers is subject to certain restrictions, which in the present case have no very obvious economic significance. This section provides a rigorous analysis of conditions under which the tax formulae (34) are indeed necessary conditions for optimality, and in particular provides economically meaningful assumptions that ensure their validity. The reader should be warned that the discussion is highly technical.

One might hope to provide a rigorous analysis by using the well-known Kuhn-Tucker theorem for differentiable (not necessarily concave) functions. This theorem requires a certain "constraint qualification" to be satisfied. Let us apply it and see how far we get. We wish to

$$\text{Maximize } V(q)$$
$$\text{subject to } g(X(q)) \leqq 0 \quad \text{and} \quad q \geqq 0,$$

where g is a (vector) production constraint such that $g(X) \leqq 0$ if, and only if, X is in G. Given that V, X, and g are differentiable, and that the Kuhn-Tucker constraint qualification is satisfied, we have the first-order conditions

$$(96)\quad V'(q^*) = \frac{\partial V}{\partial q} \leqq p \cdot \frac{\partial X}{\partial q} = p \cdot X'(q^*),$$

where $p = \lambda \cdot g'(X(q^*))$ for a vector of Lagrange multipliers λ, and is therefore a support or tangent hyperplane to G at $X(q^*)$. Since V and X are homogeneous of degree zero, $[V'(q^*)-p\cdot X'(q^*)]\cdot q^*=0$: consequently $\partial V/\partial q_i=p\cdot(\partial X/\partial q_i)$ for i such that $q_i^*>0$.

---

[4] Another case can be treated in a similar manner: that of limited government production of a good, which is also being produced privately, when government production is given away rather than being sold. Since the government production rule given above does not reduce to the first-order condition in producer prices, we would not find aggregate production efficiency for the sum of these two sources of production.

To express the first-order conditions in this form, we naturally expect to assume that V and X are continuously differentiable: to that extent, the differentiability assumptions are innocuous. The assumption that the production set can be described by a finite number of continuously differentiable inequality constraints that satisfy the constraint qualification is less satisfactory. The constraint qualification is an assumption about the functions g: one can violate it by changing the functions g without changing the actual constraint set, G. Some such assumption is required to avoid not unreasonable counter-examples, as we shall see below. But it is not at all obvious how one would check whether a particular example that failed to satisfy the constraint qualification could be put right by describing G by a better behaved set of inequalities. We should like to use a constraint qualification that depends on the properties of the set G (and X) rather than the particular functions g; and we should like the assumption to be more amenable to economic interpretation. The theorem we prove below contains such an assumption, for the case where G is convex and has an interior.

Before stating the theorem let us consider an example in which the first-order conditions are not satisfied at the optimum.

Example g. Consider the one-consumer economy. In the case shown in Figure 10, the offer curve is tangent to the production frontier at the optimum production point. As q varies, the vector $X(q)$ traces out the offer curve. Thus, holding q_2 constant, the vector $\partial X(q)/\partial q_1$ is tangent to the offer curve at $X(q^*)$. Therefore if p is the vector of producer prices, which is tangent to the production frontier at $X(q^*)$, $p\cdot\partial X(q^*)/\partial q_1=0$. The same is true for the derivatives with respect to q_2. But there is no reason why $V'(q^*)$ should be zero: therefore the above first-order conditions may not be satisfied at the optimum.

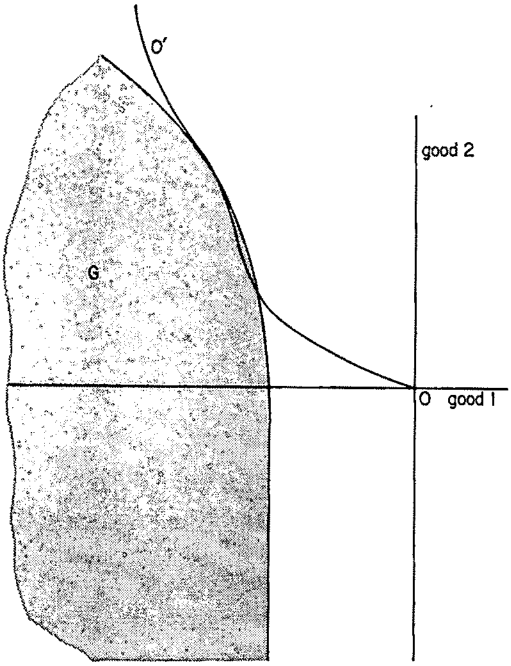


FIGURE 10

We shall make an assumption ruling out tangency between the frontier of the production set and the offer curve:

> For any p, q ($q\geqq 0$, $p\neq 0$) such that $X(q)$ is in G and $p\cdot X(q)\geqq p\cdot x$ for all x in G, $p\cdot X'(q)\geqq 0$.

The qualification takes this particular form because we also have the constraint $q\geqq 0$. Let us note that for $q>0$ the condition $p\cdot X(q)\geqq 0$ is equivalent to $p\cdot X'(q)\neq 0$, because X is homogeneous of degree zero. The qualification asserts that *for any possible competitive equilibrium (under commodity taxation) there is a consumer price change which will decrease the value of equilibrium demand, measured in producer* *prices.* By the aggregate consumer budget constraint, $q\cdot X=(p+t)\cdot X=0$. Therefore the assumption says that at any possible equilibrium point on the production frontier, it is possible to increase tax revenue. Thus the first-order conditions may not be applicable if the optimal point represents a local tax revenue maximum. Returning to example *g*, we see that $p\cdot X'=0$ at the optimum, or equivalently $\partial(t\cdot X)/\partial t=0$, although the derivatives of V are not necessarily zero there.

We now state and prove the theorem.[5]

THEOREM 5: *Assume an optimum,* (X^*, q^*) *exists; that* $V(q)$ *and* $X(q)$ *are continuously differentiable; and that* G *is convex and has a nonempty interior. Assume furthermore that there is no pair of price vectors* (p, q) *for which*

$$\begin{aligned} &X(q) \textit{ maximizes } p\cdot x \textit{ for } x \textit{ in } G,\\ &p \neq 0, \quad \textit{and} \\ &p\cdot X'(q) \geqq 0 \end{aligned} \tag{97}$$

Then there exists p^* *such that*

$$X^* \textit{ maximizes } p^*\cdot x \textit{ for } x \textit{ in } G, \quad \textit{and}$$
$$V'(q^*) \leqq p^*\cdot X'(q^*)$$

PROOF:

Let $P=\{p \mid p\cdot X^* \geqq p\cdot x, \text{ all } x \text{ in } G\}$. P is the cone of normals to G at X^*, including the zero vector. It is a nonempty, closed, convex cone.

We write V' for $V'(q^*)$ and X' for $X'(q^*)$. Consider the set

$$B = \{v \mid v \leqq p\cdot X', \text{ some } p \text{ in } P\}$$

We have to show that V' is in B. We do this by showing first, that if V' is in $\bar{B}$, the closure of B, in fact V' is in B; and then that V' must be in $\bar{B}$.

If V' is in $\bar{B}$, there exist sequences $\{v_n\}$ and $\{p_n\}$, p_n in P, such that

$$\begin{aligned} &v_n \leqq p_n\cdot X',\\ &v_n \rightarrow V' \qquad (n\rightarrow\infty) \end{aligned} \tag{98}$$

Either $\{p_n\}$ is bounded or it is not. If not, we can find a subsequence on which

$$\|p_n\| \rightarrow \infty, \qquad \frac{p_n}{\|p_n\|} \rightarrow \bar{p} \neq 0$$

Then, dividing (98) by $\|p_n\|$ and letting $n\rightarrow\infty$ on the subsequence, we obtain $\bar{p}\cdot X' \geqq 0$ while $\bar{p}$, $\neq 0$, is in P. This possibility is excluded by assumption (97). Therefore $\{p_n\}$ is bounded, and has a limit point p, in P. Equation (98) implies that $V' \leqq p\cdot X'$. The conclusion of the theorem is thus established on the assumption that V' is in $\bar{B}$.

Suppose, on the contrary, that V' is not in $\bar{B}$. We shall derive a contradiction by a sequence of lemmas.

LEMMA 5.1:

$\bar{B}$ is pointed. That is, v and $-v$ both belong to $\bar{B}$ only if $v=0$.

PROOF:

If v, $-v$ is in $\bar{B}$, we have sequences such that

$$\overset{1}{v}_n \leqq \overset{1}{p}_n\cdot X', \qquad \overset{2}{v}_n \leqq \overset{2}{p}_n\cdot X', \tag{99}$$

$$\overset{1}{v}_n \rightarrow v, \qquad \overset{2}{v}_n \rightarrow -v \tag{100}$$

If $v\neq 0$, it cannot be the case that p_n^1 and p_n^2 both tend to zero. Suppose, for example, p_n^1 does not, and take a subsequence on which

---

[5] It should be noticed that when the constrained optimum is (locally) an unconstrained maximum, the producer prices satisfying the theorem are zero. This happens if optimal production is in the interior of the production set and may happen if it is on the frontier. The theorem can be weakened in a complicated manner by replacing the nontangency qualification by two conditions. One is an analog of the Kuhn-Tucker Constraint Qualification providing for the existence of an arc in the attainable set. The other use of nontangency occurs when V' is in $\bar{B}$ but not in B. If it is assumed that when there is tangency, the cone of normals is polyhedral, B will be closed. The Kuhn-Tucker theorem is then a special case of the weakened version of theorem 5 when G is the nonnegative orthant. The Kuhn-Tucker theorem is very much easier to prove, however.

$$\|p_n^1\| \to \pi_1 \leqq \infty,$$

$$p_n^1/\|p_n^1\| \to p^1, \neq 0$$

If $p_n^1+p_n^2 \to 0$, $p_n^2/\|p_n^1\| \to -p^1$, and therefore $-p^1$ is in P. This is impossible, since, G having a nonempty interior, P is pointed. (If p, $-p$ are in P, $p\cdot x$ is constant for x in G, but a hyperplane has no interior.) We can therefore take a subsequence on which

$$\|p_n^1 + p_n^2\| \to \pi, \qquad 0 < \pi \leqq \infty,$$

$$\frac{p_n^1 + p_n^2}{\|p_n^1 +_n^2 p\|} \to p, \neq 0, \in P$$

From (99) (adding and dividing by $\|p_n^1+p_n^2\|$) and (100), we now have

$$p\cdot X' \geqq \text{Lim}\, \frac{v_n^1 + v_n^2}{\|p_n^1 + p_n^2\|} = 0 \tag{101}$$

This contradicts (97), since p is in P and $p \neq 0$, and thereby establishes the lemma.

LEMMA 5.2: *If C is a pointed, closed, convex cone, there exists a vector p such that for all non-zero z in C, $p\cdot z<0$.*

PROOF:

By the duality theorem for convex cones $C^{++}=C$, where C^+ is the dual cone, $\{p \mid p\cdot z \leqq 0,\ z \text{ is in } C\}$. Clearly, if C^+ is pointed, C has a nonempty interior: for if interior C is empty, $p\cdot z=0$ for some nonzero p and all z in C, and then p and $-p$ both belong to C^+. Under the assumptions of the theorem, C is closed and pointed. Therefore C^{++} is pointed, and C^+ has an interior point p.

$$p\cdot z < 0 \quad (\text{all nonzero } z \text{ in } C)$$

Otherwise, if $p\cdot z=0$, we can easily find a sequence $\{p_n\}$ on which $p_n \to p$ and $p_n\cdot z>0$, so that p_n is not in C^+.

LEMMA 5.3: *If V' is not in $\overline{B}$, there exists r such that*

$$V'\cdot r > 0 \tag{102}$$

$$v\cdot r < 0 \qquad (v \in B) \tag{103}$$

PROOF:

The closed convex cone $\overline{B}+\{\lambda V' \mid \lambda \leqq 0\}$ is pointed. Thus there exists an r such that

$$v\cdot r + \lambda V'\cdot r < 0$$

$$(v \in \overline{B},\ \lambda \leqq 0,\ v, \lambda \text{ not both zero})$$

Putting $v=0$ and $\lambda=-1$ we obtain (102); putting $\lambda=0$ we obtain (103).

LEMMA 5.4: *Let r be a vector satisfying (102) and (103). For some $\delta>0$,*

$$X(q^* + \theta r) \in G \quad (0 \leqq \theta \leqq \delta) \tag{104}$$

PROOF:

Assume not. Then for some sequence $\{\theta_n\}$, $\theta_n>0$, $\theta_n \to 0$,

$$X(q^* + \theta_n r) \notin G$$

Since G is convex, this implies that

$$X(q^*) + \frac{\lambda}{\theta_n}\left[X(q^* + \theta_n r) - X(q^*)\right] \notin G$$

for $\lambda \geqq \theta_n$. Letting $n \to \infty$, we deduce, for any $\lambda>0$, that

$$X(q^*) + \lambda X'\cdot r = \lim_{n\to\infty}\left[X(q^*) + \lambda\frac{X(q^* + \theta_n r) - X(q^*)}{\theta_n}\right]$$

is not in the interior of G. It follows that the half-line $\{X(q^*)+\lambda X'\cdot r \mid \lambda>0\}$ can be separated from the interior of G by a hyperplane with normal $p \neq 0$:

$$p\cdot X(q^*) + \lambda p\cdot X'\cdot r \geqq p\cdot x$$

$$(\lambda > 0,\ x \in \text{Int } G)$$

Letting $\lambda \to 0$ we have $p \in P$. Letting $x \to X^*$ we have

$$p\cdot X'\cdot r \geqq 0,$$

which contradicts (103) since $p\cdot X'$ is in B. The lemma is proved.

Since q^* is optimal, (104) implies that

$$V(q^* + \theta r) \leqq V(q^*) \qquad (0 \leqq \theta \leqq \delta)$$

Therefore,

$$\begin{aligned} V' \cdot r &= \operatorname*{Lim}_{\theta \to 0} \frac{1}{\theta} [V(q^* + \theta r) - V(q^*)] \\ &\leqq 0 \end{aligned}$$

This, however, contradicts (102). The hypothesis of Lemma 5.3, that $V' \in \bar{B}$, is therefore false. The proof of the theorem is thus complete.

In reaching our results that the first-order conditions for optimum taxes (96) hold in general, we have assumed that the production set, G, is convex. But one common argument for government control of production is nonconvexity of the production set. This is not a question we are primarily concerned with in this paper. However, some extensions of the theorem do hold. As an example, assume the frontier of G is differentiable at X^*, so that p can be uniquely defined as the normal at X^* and that G is not thin in the neighborhood of X^*—i.e., there exists a ball with center on the normal through X^*, contained in G and containing X^*. Applying the theorem to this ball we get the validity of the first-order conditions (96) using the producer prices defined by the normal.

As in general welfare economics, two uniqueness problems may arise when considering the application of the first-order conditions to achieve an optimum. In the first place, there may be more than one pair of price vectors, (p, q), that satisfy the first-order conditions and allow markets to be cleared. This is similar to the problem that arises when we attempt to define optimum production and distribution by first-order conditions in the presence of a non-convex production set. It is noteworthy that, if lump sum transfers are excluded as a feasible policy, this problem may arise even when the production set is convex. There is no reason why the demand functions should have any of the nice convexity properties which ensure that first-order conditions imply global maximization. Only in particular cases, such as that discussed in footnote 2 above (where rigorous argument is possible without appeal to theorem 5), will the first-order conditions lead to a unique solution.

The second problem is that the tax policies one might like to employ may not uniquely determine the behavior of the system. The lump sum redistribution of wealth required in standard welfare economics does not carry with it any guarantee that the desired competitive equilibrium is the unique one consistent with the optimal wealth distribution (although if the wrong equilibrium is achieved, this should be easily noticed). Similarly, in the present case, if we employ taxes rather than consumer prices as the government control variables, the equilibrium of the economy may not be unique.[6] But if consumer prices are used as the control variables—and why not?—the demand functions give us a unique equilibrium position, so long as preferences are strictly convex.

XI. Concluding Remarks

Welfare economics has usually been concerned with characterizing the best of attainable worlds, accepting only the basic technological constraints. As economists have been aware, the omitted constraints on communication, calculation, and administration of an economy (not to mention political constraints) limit the direct applicability of the implications of this theory to policy problems, although great insight into these problems has certainly been acquired. We have not attempted to come directly to grips with the problem of incorporating these complications into economic theory. Instead, we have explored the implications of viewing these constraints as limits on the set of policy tools that can be applied. There are many sets of policy tools which might be examined in this way. Specifically, we have assumed that the policy tools available to the government include commodity taxation (and subsidization) to any extent. For these tools we have derived the rules for optimal tax policy and have shown the desirability of aggregate production efficiency, in the presence of optimal taxation. We have also considered expansion of the set of policy tools in such a way that we continue to have the condition that production decisions do not change the class of possible budget constraints. For example, this condition is still preserved when one includes poll taxes, progressive income taxation, regional differences in taxation, taxation on transactions between consumers, and most kinds of rationing. This type of expansion of the set of policy tools does not alter the desirability of production efficiency, nor does it alter the conditions for the optimal commodity tax structure, although in general the tax rates themselves will change. We have, unfortunately, ignored the cost of administering taxes. Presumably optimization by means of sets of policy tools that do not, because the cost of administration, include the full scope of commodity taxation, will not lead to the same conclusions.

---

[6] For a discussion of multiple equilibria in a related problem, see E. Foster and H. Sonnenschein.

Let us briefly consider the type of policy implications that are raised by our analysis. In the context of a planned economy our analysis implies the desirability of using a single price vector in all production decisions, although these prices will, in general, differ from the prices at which commodities are sold to consumers.

As an application of this analysis to a mixed economy, let us briefly examine the discussion of a proper criterion for public investment decisions. As has been widely noted, there are considerable differences in western economies between the intertemporal marginal rates of transformation and substitution. This has been the basis of analyses leading to investment criteria which would imply aggregate production inefficiency because they employ an interest rate for determining the margins of public production which differs from the private marginal rate of transformation. One argument used against these criteria is that the government, recognizing the divergence between rates of transformation and substitution, should use its power to achieve the full Pareto optimum, bringing these rates into equality. When this is done, the single interest rate then existing will be the appropriate rate to use in public investment decisions. We begin by presuming that the government does not have the power to achieve any Pareto optimum that it chooses. Then from the maximization of a social welfare function, we argued that the government will, in general, prefer one of the non-Pareto optima to the Pareto optima, if any, that can be achieved. At the constrained optimum, which is the social welfare function maximizing position of the economy for the available policy tools, we saw that the economy will still be characterized by a divergence between marginal rates of substitution and transformation, not just intertemporally, but also elsewhere, e.g., in the choice between leisure and goods. However, we concluded that in this situation we desired aggregate production efficiency. This implies the use of interest rates for public investment decisions which equate public and private marginal rates of transformation.

We have obtained the first-order conditions for public production, but we have not considered the correct method of evaluating indivisible investments. This is one problem that deserves examination. In examining the optimal tax structure, we have briefly considered the tax rates implied by particular utility functions. This analysis should be extended to more general and more interesting sets of consumers. Further, we have not examined in any detail the uniqueness and stability of equilibrium, that is, the question whether there are means of achieving in practice an equilibrium which is close to the optimum.

Finally, we would like to emphasize the assumptions which seem to us most seriously to limit the applications of this theory.[7] We have assumed no costs of tax administration and no tax evasion. And we have assumed constant-returns-to-scale and price-taking, profit-maximizing behavior in private production. Pure profits (or losses) associated with the violation of these assumptions imply that private production decisions directly influence social welfare by affecting household incomes. In such a case, it would presumably be desirable to add a profits tax to the set of policy instruments. Nevertheless, aggregate production efficiency would no longer be desirable in general; although it may be possible to get close to the optimum with efficient production if pure profits are small. We hope, nevertheless, that the methods and results of this paper have shown that economic analysis need not depend on the simplifying, but unrealistic, assumption that the perfect capital levy has taken place.[8]

REFERENCES

W. J. Corlett and D. C. Hague, "Complementarity and the Excess Burden of Taxation," *Rev. Econ. Stud.*, 1953, *21*, No. 1, 21–30.

E. Foster and H. Sonnenschein, "Price Distortion and Economic Welfare," *Econometrica*, Mar. 1970, *38*, 281–97.

H. Kuhn and A. Tucker, "Nonlinear Programming," in J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley 1951.

J. A. Mirrlees, "An Exploration in the Theory of Optimum Income Taxation," *Rev. Econ. Stud.*, Apr. 1971, *38*, forthcoming.

C. C. Morrison, "Marginal Cost Pricing and the Theory of Second Best," *Western Econ. J.*, June 1969, *7*, 145–52.

P. A. Samuelson, "Memorandum for U.S. Treasury, 1951," unpublished.

———, "The Pure Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.

---

[7] These assumptions are viewed in the context of equilibrium theory. There is no need here to go into the limitations inherent in current equilibrium theory.

[8] A recent paper by Clarence Morrison also deals with marginal cost pricing as a special case of optimal pricing.

A Short-Term Econometric Model of Textile Industries

By Roger LeRoy Miller*

Short-term models have been devised for the purpose of answering both general and specific questions. These models have ranged from simple descriptions of an industry or an economy to specific estimates of certain parameters, such as output-factor elasticities or adjustment rates. The results presented here describe, in a somewhat unusual manner, two textile industries. The estimates of output-factor elasticities differ from those obtained by other investigators using single-equation estimation techniques. The model consists of a dichotomized (peak, off-peak) system of equations that attempts to capture relevant differences in decision making by firms during peak and off-peak periods in the production cycle. Investment in labor and in product inventories is considered, along with estimation of a production function and a demand-for-hours equation.

When an industry such as textiles is considered, production data clearly show definite seasonal fluctuations. In the analysis of decisions such as hiring and firing of workers, no adequate theory can ignore these (imperfectly) known fluctuations. During peak production periods most firms are faced with capacity constraints which do not exist during off-peak periods in the production cycle. During the latter the level of productive (efficient) employment of the hired work force may be considered as determined by "desired" capital-labor proportions of previous peak periods. That is, given the average full-capacity capital-labor ratio (presumed optimum)[1] and the rate of production during any off-peak period, the employment decision of the firm is not the total number of employees to hire, but rather the number to retain in excess of that dictated by the "optimal" capital-labor ratio and the rate of production. A prime decision is thus how much to invest in inventories of labor (reserve labor force) during seasonal partial-capacity production periods. Kenneth Arrow et al. (p. 18) recognized the need for an inventory approach to the demand for labor and Hollis Chenery considered similar "over-investment" behavior of firms as related to physical capacity. I extend this approach by focusing on inventories of labor and their interaction with other variables such as product inventories and production-worker hours.

Considering the relation between production (Q), sales (S), and finished-goods inventories (I),[2] a firm can hold production constant (never needing inventories of labor) while satisfying all changes in sales out of product inventories. Or it can vary production rates to match sales fluctuations. Or it can use a combination of both. If costs of inventory adjustment (C_I) are "high" relative to those of pro-

---

* University of Washington, Seattle. Many thanks are extended to Zvi Griliches for his numerous comments and criticisms and to an unnamed referee. Yoram Barzel, Tom Borcherding, Bruce Gardner, John Floyd, Leroy Hushak, Potluri Rao, Sherwin Rosen, Gerald Scully, Tom Simpson, and Arnold Zellner made helpful comments on an earlier draft of this paper.

[1] It is understood that this is an oversimplification; a firm producing steadily at peak output will employ fewer men and machines than it would have done at a seasonal peak.

[2] Changes in finished-goods inventories, $\Delta I_t = Q_t - S_t$.

duction rate changes, it may choose to vary production rates; then it must decide how to vary its stock of inputs relative to its production rate. Whenever the cost of of hiring, firing, and training exceed those of retention of inventories of labor (Z), the firm will hold the latter. Numerous authors have discussed the potential saving in turnover costs via the holding of a "reserve" labor force. (See, for example, Charles Holt et al. and Sumner Slichter.)

Walter Oi clearly demonstrated empirically the existence of labor as a "quasi-fixed" factor, as did Sherwin Rosen (1968); and Gary Becker theoretically treated this problem in human capital investment. Holt and M. H. David realized that the desired work force depends, *inter alia*, on finished goods' inventory and on forecasts of future sales, so that " . . . new vacancies need to be created in anticipation of future needs, ..." If these labor inventory vacancies were not filled, the firm " . . . would then face the following costly alternatives: working overtime, losing sales, or reducing [product] inventories to a level lower than desired" (p. 82). Holt (1969) implicitly suggested the existence of a reserve labor force when remarking that in response to a production increase " . . . labor productivity of the existing work force then increases by about 1 percent through *reducing slack time*, etc." (p. 137, emphasis added).

In what follows, I integrate the reserve labor theory into a system consisting of the following functions: production, peak-period demand for labor, inventories of labor, peak and off-peak demand for production worker hours, and peak and off-peak demand for finished-goods inventories.

I. The Production Function

A production relation should relate actual quantities produced to actual (as opposed to measured) quantitites utilized in the production process. The following Cobb-Douglas function is assumed:[3]

$$Q = A(E'H)^b K^c e^{p\mathrm{T}} \tag{1}$$

where Q = quantity produced, E' = the number of technically effective employees, H = average hours worked per effective employee, K = total machine hours, T = a smoothly ascending time variable, and A, b, c, and p = parameters to be estimated.

Assume that E' is identically equal to measured labor E, whenever the firm is operating at full capacity. That is, during peak periods inventories of labor are nonexistent, E' becomes E in (1). The difference between E' and E is inventories of labor.

II. The Demand for Inventories of Labor

Given that firms desire a reserve labor force during slack production periods, what determines the desired stock? The firm typically has some idea of anticipated future changes in sales, especially if there are recurrent seasonals. For anticipated increases, there are two ways to assure that excessive backlogs do not pile up. One is to maintain larger levels of finished-goods inventories from which sales in excess of production can be met; the other is to maintain excess capacity in terms of men and machines so that the rate of production can be increased to match anticipated increased demand. It is thus obvious that, *ceteris paribus*, inventories of labor (and concomitant excess physical capacity) are substitutes for inventories of finished goods. That is, again *ceteris paribus*, desired reserve labor (Z^d) should be negatively related to finished goods inventories. This hypothesized substitution exists merely because these two inventories compete for the same purpose—to reduce the cost to the firm of backlogs.

---

[3] For each short-run period, I assume negligible substitution between efficient units of labor and capital services, although not in the long run. The Cobb-Douglas function, with unitary substitution elasticity, is estimated with data from peak periods, thus allowing for long-run substitution (since most observations are four quarters apart).

It must be mentioned also that for any known sales course, a change in relative holding costs will induce a shift to a higher optimum level of the relatively cheaper inventory. For example, given the cost of holding I, any change in wages should result in a change in desired reserve labor. Naturally if there is little or no factor substitution, the wage rate, W, cannot affect the effective capital-labor mix. If, however, the costs of labor-stock adjustment such as recruitment, selection, training, etc., do not rise proportionally with the wage rate, then—given product-inventory holding costs—a rise in wages should induce a decrease in Z^d. Since I assume that the reserve labor force is not worked overtime, the relevant price here is the straight-time wage.[4]

It is true that one can think of situations where desired inventories will all move in the same direction. If a much larger permanent increase in sales is anticipated, the present value of the marginal revenue for holding reserve labor and for holding goods' inventories will exceed their marginal costs; hence both will be increased. Or, if the production cycle is perfectly in phase with the sales cycle but with a smaller amplitude, the corresponding inventory cycle will be positively related to the cycle of reserve labor. Since nothing in this model explicitly yields any indication of the timing of production relative to sales, it may be that the hypothesized negative relation between the two types of inventories will not show up. The problem is further complicated by the lack of explicit cost functions for the two inventories. Suffice it to say that the relative holding costs are changing and that if the data generate a negative relation between labor and finished-goods inventories (Z^d and I) and between labor inventories and wages (Z^d and W), then in some sense the idea of inventory substitution can be accepted.

The demand function for Z^d should fall out of the relative cost function, which would include some measure of expected future sales. Without undertaking this task, I merely specify that Z^d is a multiplicative function of a few key variables for which data should be available:

$$Z^d = \Lambda(S^*)^\delta I^r W^\phi C_I^\xi \mathrm{T}^\pi \tag{2}$$

Desired labor reserve depends on a constant (Λ), anticipated future sales (S^*), product inventories (I), straight-time wages W, product-inventory adjustment costs (C_I), and a trend (T) added to take account of changes in secular demand.

Since I am assuming that firms do not hold inventories of labor during peak production periods, it is necessary to formulate an equation for peak-period labor demand.

III. Derived Demand for Labor During Peak Periods

Using an approach developed by Robert Ball and E. B. A. St. Cyr, assume a simple cost function:

$$C = W^*(EH) + F \tag{3}$$

where C = total direct costs net of materials, fuel, and variable capital user costs; W^* = the effective wage rate per man-hour; and F = fixed costs.

We note that W^*, being the effective pay per productive hour worked, will be "high" at lower hours per week worked and will gradually drop to a minimum when overtime rates are instituted, then will rise again. We can use a quadratic approximation to represent this relation between W^* and H:

---

[4] Note that though inventories of labor are measured in numbers of men, this may be an oversimplification since the reserve labor force may be composed of a fraction of every man on the payroll during off-peak production periods.

$$W^* = g_0 - g_1 H + g_2 H^2 \tag{4}$$

For peak periods W^* will be a combination of straight and overtime pay and will thus be to the right of the minimum point mentioned above.

Putting (4) into (3) and adding the production function (1) constraint, the Lagrangian L is formed:

$$\begin{aligned} L = {} & g_0 EH - g_1 EH^2 + g_2 EH^3 + F \\ & - \lambda \left[Q - A(EH)^b K^c e^{p\mathrm{T}} \right] \end{aligned} \tag{5}$$

Differentiating L with respect to E and H, setting the results equal to zero, and solving, gives static-equilibrium solutions for E and H:[5]

$$E^* = \left(\frac{2g_2}{g_1} \right) A^{-1/b} Q^{1/b} K^{-c/b} e^{-p\mathrm{T}/b} \tag{6}$$

$$H^* = \left(\frac{g_1}{2g_2} \right) \tag{7}$$

Desired level of employment is a multiplicative function of a constant, output, capital, and time; whereas desired static level of hours is a constant.

I assume that due to adjustment costs there will be lags in the firm's peak-period response to changes in sales, hence in production and employment. Thus the standard partial adjustment mechanism is used with one *ad hoc* addition, a factor that takes into account how far actual hours vary from equilibrium hours (H^*). (E_{t-1} is last-period measured employment even if it was off-peak.)

$$\left(\frac{E_t}{E_{t-1}} \right) = \left(\frac{E^*}{E_{t-1}} \right)^q \left(\frac{H^*}{H} \right)^a \tag{8}$$

Substituting (8) into (6), and using (7), the peak-period labor demand function is:[6]

$$\begin{aligned} E = {} & \left(\frac{2g_2}{g_1} \right)^{q-a} A^{-q/b} Q^{q/b} K^{-cq/b} \\ & \cdot E_{t-1}^{(1-q)} H^{-a} e^{-qp\mathrm{T}/b} \end{aligned} \tag{9}$$

Now actual peak labor demanded is a function of output, capital, lagged employment, production-worker hours, and time.

Desired hours (H^*) are shown to be a constant (7). This static solution needs modification for use in a short-run model.

IV. Demand for Production-Worker Hours

Hours worked provide the easiest and fastest variable for a firm to alter during transitory periods of change in production. Since little more than casual observation has been put forward on the demand for hours,[7] I apply in an *ad hoc* manner a modified partial adjustment mechanism to (7) to obtain the equation for peak-period hours demand:[8]

$$H = \left(\frac{g_1}{2g_2} \right)^j H_{t-1}^{(1-j)} Q_t^l Q_{t-1}^{-l} \tag{10}$$

Actual peak-period hours demanded are a multiplicative function of lagged hours and transitory changes in production Q_t/Q_{t-1}.

It is probable that a firm finding itself with large inventories of labor will wish to reduce paid-for hours faster for off-peak periods than otherwise, *ceteris paribus*. Thus we should find a negative relation between reserve labor, Z, and total hours p14417

---

[5] Second-order conditions are easily verified. See Ball and St. Cyr and N. J. Ireland and D. J. Smyth for other derivations.

[6] Stability for (9) requires $0 \leq q \leq 1$ and $a > 0$.

[7] For an exception see Yoram Barzel and especially Rosen (1968, 1969).

[8] Stability requires that if $bl < 1$, then $j \leq (1 - bl)$ and if $bl > 1$, then $j \geq (1 - bl)$. As an unnamed referee has pointed out, since changes in H affect labor holding costs a formal link should exist between inventory and employment functions and the demand-for-hours equation. Equations (10) and (11) do take account of this.

paid-for per man per week. Therefore the reserve labor variable[9] is added to the equation (10) for peak-period hours demand to yield the off-peak period equation:

$$H = \left(\frac{g_1}{2g_2}\right)^j H_{t-1}^{(1-j)} Q_t^l Q_{t-1}^{-l} Z^m \tag{11}$$

V. The Demand for Inventory Investment

The theoretical results obtained by M. C. Lovell are used here except for two modifications. All identities have been put into ratio form (which holds only approximately) in order to make the demand equation multiplicative. Also, the reserve-labor variable was added throughout in accordance with my hypothesis that Z^d and I are substitutes. For brevity, I will not go through Lovell's theory[10] but will present only the resultant inventory-investment equation:

$$\frac{I_t}{I_{t-1}} = T^h I_{t-1}^k (S_{t-1}^t)^n S_t^r (S_{t-1}^{t+1})^u \cdot (S_t^{t+1})^s (Z_{t-1}^t)^v Z_t^w \tag{12}$$

Variables subscripted and superscripted are anticipated ones: i.e., S_{t-1}^t is t'th period sales as predicted in period $t-1$.[11]

Equation (12) indicates that inventory investment I_t/I_{t-1} is a multiplicative function of lagged inventories I_{t-1}, t'th period sales as predicted using information up to period $t-1$ (S_{t-1}^t), actual t'th period sales (S_t), $t+1$ period sales as predicted in period $t-1$ (S_{t-1}^{t+1}), $t+1$ period sales as predicted in period t (S_t^{t+1}), predicted labor reserves (Z_{t-1}^t), and actual labor reserves (Z_t). The exponent terms h, K, n, r, u, s, v, and w, are to be estimated.

VI. Data and Estimation

Two textile industries are covered in this section: woolen weaving (221) and cotton weaving (223). The Bureau of Census publishes short-term data for these industries on capital services (machine hours), production, and inventories. Data on production workers and their hours are reported by the Bureau of Labor Statistics. Three variables had to be constructed: straight-time hourly wages (*HW*), expected sales (S^*), and inventories of labor (Z).

The wage rate is obtained by using data on average hourly earnings (*AHE*), average production worker weekly hours (*AWH*), and average overtime hours (*AOTH*).[12]

Expected sales are constructed with information on past sales by using a weighted moving-average technique developed by Holt et al. which utilizes seasonal and trend information. Since three weights needed specifying, it was decided to iterate on a three-parameter space until the constructed future sales series correlated most highly with actual sales, *ex post*. That is, over 200 predicted series were computed using different combinations of weights, and the "best" series determined the actual weights used.

A measure of Z was obtained by assuming that capital is utilized most efficiently when average hours per machine per week are at peak (full capacity). I assume also that the firm has some optimum ratio of men to machines that can be observed during these peak periods.[13] By plotting ma-

---

[9] As will be seen later, Z is an endogenous variable used as a proxy for Z^d in the simultaneous system.

[10] My full analysis will be furnished on request.

[11] Notice that (12) holds in its present form only for off-peak periods, since it equals zero when $Z=0$. To avoid this, assume that the reserve-labor variable is Z/Z^*, where Z^* is some long-run desired level, and that during peak period $Z/Z^*=1$. Then Z^* goes into the constant and the peak-period equation does not include the variable.

[12] Assuming that overtime pay equals 150 percent of straight pay, then:

$$HW = \frac{(AWH)\cdot(AHE)}{(1.5)\ (AOTH) + (AWH - AOTH)}$$

[13] Some firms have maintained that their most

chine hours on a time-series graph, capital-use peaks could readily be identified. (These are the peak periods for estimation of the full system.) The ratio of men to machines at these peaks was then computed and the result called "desired" ratios. A three-part moving average of the ratios was then calculated. For all off-peak periods the effective, or efficient, labor force, E_t, was obtained by:

$$(13)\quad E'_t = \frac{\text{observed machines in operation in period t}}{\text{average desired machine-men ratio}}$$

The reserve labor force was then obtained by:

$$(14)\qquad Z_t = E_t(\text{observed}) - E'_t$$

This method allows for increases in labor productivity *pari passu* while still utilizing the assumption of negligible short-run capital-labor substitution. It is empirically accurate if the firm correctly anticipates peaks and has constant returns.

The final estimating form of the dichotomized system in logs with log-normal error terms added is:

Peak Period System

$$(1')\quad \log Q = \text{const.} + b \log (EH) - c \log K + pT + u_1$$

$$(9')\quad \log E = \text{const.} + \frac{q(1-c)}{b} \log Q + (1-q) \log E_{t-1} - a \log H - \frac{cp}{b} T + u_{9'}$$

$$(10)\quad \log H = \text{const.} + (1-j) \log H_{t-1} + l \log Q - l \log Q_{t-1} + u_{10}$$

$$(12')\quad \log \left(\frac{I_t}{I_{t-1}}\right) = \text{const.} - k \log I_{t-1} + r' \log Q + s \log S_t^{t+1} + u_{12'}$$

Off-Peak Period System:

$$(2')\quad \log Z = \text{const.} + \delta \log S_t^{t-1} + \tau \log I + \phi \log W + \pi T + u_{2'}$$

$$(11)\quad \log H = \text{const.} + (1-j) \log H_{t-1} + l \log Q - l \log Q_{t-1} + m \log Z + u_{11}$$

$$(12'')\quad \log \left(\frac{I_t}{I_{t-1}}\right) = \text{const.} - k \log I_{t-1} + s \log S_t^{t+1} + w \log Z_t + r' \log Q_t + u_{12''}$$

The estimating equations differ from those presented earlier because of certain data limitations and problems:

(1) *Log K* was absorbed into *log Q* in (9) because of its almost perfect collinearity with the latter.

(2) High collinearity between the pairs S^t_{t-1}, S_t, and S^{t+1}_{t-1}, S^{t+1}_t forced me to delete the first of each pair from the inventory equation (12). The identity $\Delta I_t = Q_t - S_t$ (in ratio form) was used to eliminate S_t from (12). For simplicity it was assumed that $Z^t_{t-1} \cong Z_t$.

(3) Data were unavailable for constructing C_I in (2). It is assumed, though, that C_I varied little in relation to wages (W), so that a negative coefficient for *log W* in (2) should still be obtained.

(4) Notice also that (2) is in terms of Z^d whereas (2') contains Z. I argued that Z^d is most importantly a function of expected sales, so that if computed expected sales are a good predictor of future actual sales, the use of Z will not introduce measurement error. Since my computed expected sales series were exceedingly well correlated with actual future sales, I consider Z an acceptable proxy for Z^d.

For the peak system, Q, E, H, and I_t/I_{t-1} are endogenous: the equations are overidentified. For the off-peak system, Z, I_t/I_{t-1}, and H are endogenous (Q being technically exogenous because of no capacity constraints). These equations are also overidentified.

---

efficient rate of production is at less than 100 percent capacity.

TABLE 1—THREE-STAGE LEAST SQUARES RESULTS, QUARTERLY, PEAK PERIODS

Equation	Independent Variable	Woolens		Cottons	
		Coefficient Estimate	Coefficient ÷ St. Error	Coefficient Estimate	Coefficient ÷ St. Error
Production	1	−2.443	−0.932	−1.288	−0.606
	$log\ (EH)$	0.655	2.233	0.427	3.466
	$log\ K$	0.442	2.002	0.635	3.781
	T	0.004	2.645	0.001	1.231
Demand for Labor	1	1.391	3.333	−1.978	−0.743
	$log\ Q$	0.451	7.989	0.465	5.491
	$log\ H$	−0.018	−0.656	−0.010	−1.149
	$log\ E_{t-1}$	0.418	3.236	0.477	0.836
	T	−0.002	−2.011	−0.001	−1.194
Demand for Hours	1	1.623	1.721	4.904	3.108
	$log\ Q$	0.101	1.994	0.977	10.288
	$log\ Q_{t-1}$	−0.099	−2.119	−1.043	−9.355
	$log\ H_{t-1}$	0.744	6.762	0.650	3.510
Inventories	1	−3.776	−0.756	−1.281	−0.311
	$log\ I_{t-1}$	−0.667	−2.123	−0.124	−1.792
	$log\ Q$	0.811	2.468	0.233	1.249
	$log\ S_{t-1}^{t+1}$	0.155	2.657	0.263	2.198

TABLE 2—THREE-STAGE LEAST SQUARES RESULTS, QUARTERY, OFF-PEAK PERIODS

Equation	Independent Variable	Woolens		Cottons	
		Coefficient Estimate	Coefficient ÷ St. Error	Coefficient Estimate	Coefficient ÷ St. Error
Reserve Labor	1	0.984	2.369	0.997	1.410
	$log\ I$	−0.401	−8.488	−0.321	−6.817
	$log\ W$	−0.146	−4.777	−0.100	−1.989
	$log\ S_{t-1}^{t+1}$	0.374	2.001	0.349	3.687
	T	−0.0001	−1.114	−0.0002	−1.043
Demand for Hours	1	1.994	2.001	3.001	1.483
	$log\ Q$	0.099	2.220	0.914	11.466
	$log\ Q_{t-1}$	−0.116	−1.449	−0.989	−7.371
	$log\ H_{t-1}$	0.668	2.798	0.569	1.982
	$log\ Z$	−0.142	−1.398	−0.055	−1.117
Inventories	1	−4.987	−2.114	−0.993	−0.009
	$log\ I_{t-1}$	−0.565	−6.428	−0.201	−2.444
	$log\ Q$	0.771	5.976	0.276	2.119
	$log\ S_t^{t+1}$	0.200	1.884	0.318	3.437
	$log\ Z$	−0.605	−4.588	−0.115	−8.457

Tables 1 and 2 give estimation results using unrestricted three-stage least squares with quarterly data, 1960 to 1967.

VII. Analysis of Empirical Results

All regression coefficients had the theoretically correct signs, although several were not statistically "significant" at conventional levels. The coefficient of I in the demand for labor reserve was indeed negative, thus indicating that changes in relative costs during the sample period did elicit a substitution. The coefficient of the only included cost variable, W, was negative as expected. Both industries exhibited 9 percent increasing returns, but a null hypothesis of constant returns could not be rejected at conventional significance levels using a *t*-test.

Much recent empirical work (cited below) on implied output-labor elasticities was consistent with increasing returns to labor alone, a proposition that standard theory does not readily accommodate. My estimates of implied production function elasticities were consistent with decreasing returns to labor alone.

It is interesting to see whether the implied output-labor elasticities from the equation for peak labor-demand conformed to those of the production function. That is, were the cross-equation restrictions satisfied? The direct-output labor elasticity b from the peak-period production function was 0.655 for woolens and 0.427 for cottons. Notice, though, that b appears as part of the coefficient of $\log Q$ in equation (9′). Since (9′) represents a difference equation, it is possible to obtain a long-run solution in terms of $\log Q$ and then to obtain the implied estimate of b.

The solution obtained is:

$$\log E = f\left[\left(\frac{1-c}{b}\right)\log Q, \ldots\right] \tag{15}$$

For woolens, inserting $c=0.442$ yields $b=0.765$. For cottons, inserting $c=0.635$ yields $b=0.785$. The first value is less than 10 percent off from the direct production

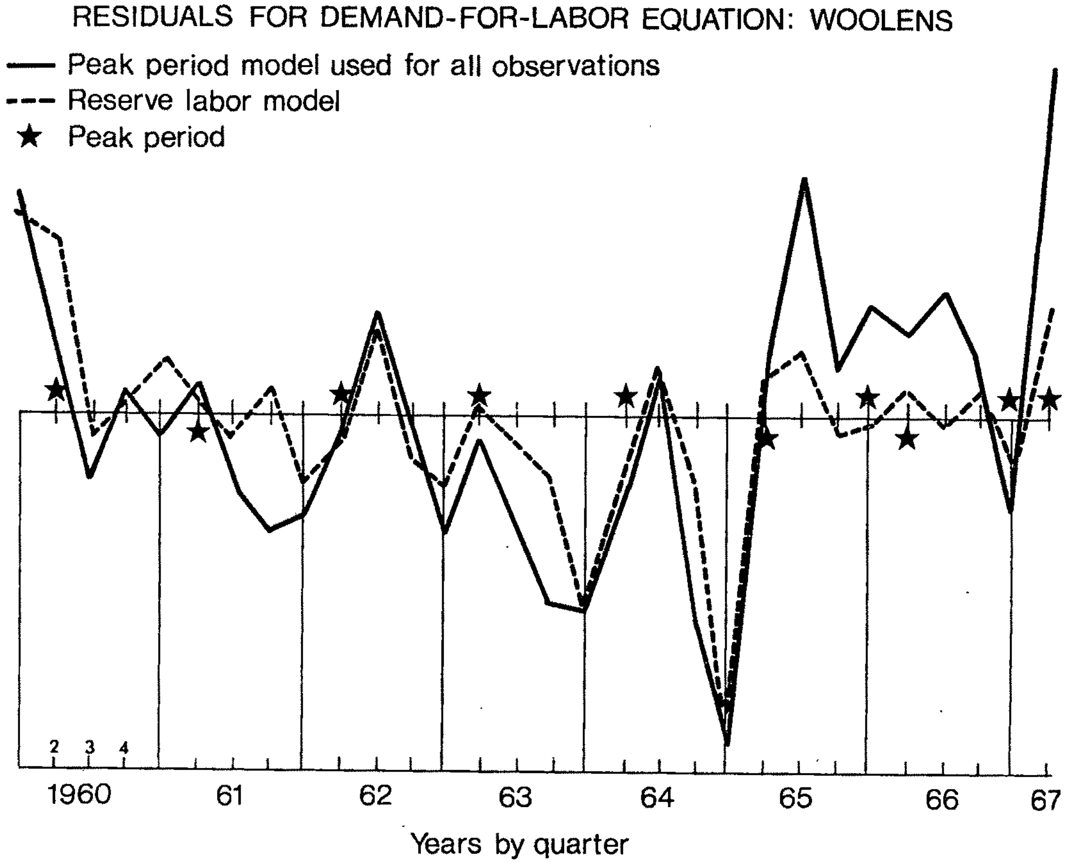


Figure 1

function estimate, but the second is substantially larger, although still less than unity. Such a small discrepancy for woolens indicates that a non-linear restriction was unnecessary during the estimation to obtain consistent cross-equation estimates of the output-labor elasticities. The same cannot be said for cottons. Alternatively, we can ignore the coefficient of $\ln H$ in (9′) and solve the resultant simple-difference equation for the implied coefficient of output: 0.77 for woolens; 0.89 for cottons.

We notice also that for the demand-for-hours equation, the magnitude of the coefficient of $\log Q$ is almost always approximately the same as that for $\log Q_{t-1}$.

Furthermore, the coefficients were always opposite in sign, thus indicating that the hours did in fact vary in response to transitory changes in output Q/Q_{t-1}, as postulated in Section IV. What is striking, though, is the large difference in these responses across industries; whereas a 1 percent change in Q/Q_{t-1} elicits about a 1 percent change in hours worked in the cotton-weaving industry, it elicits a mere one-tenth of 1 percent change in the woolen industry. This should be ample warning to researchers who attempt to use data on a more aggregated level than was done here.

Concerning the reserve labor variable, in the off-peak systems the coefficients of $\log Z$ were negative in the equations for demand for hours and demand for inventories, although not quite "significant" for the former. Moreover, in the reserve labor equation proper, the coefficient of $\log I$ was very highly significant and negative, thus lending more support to the proposi-

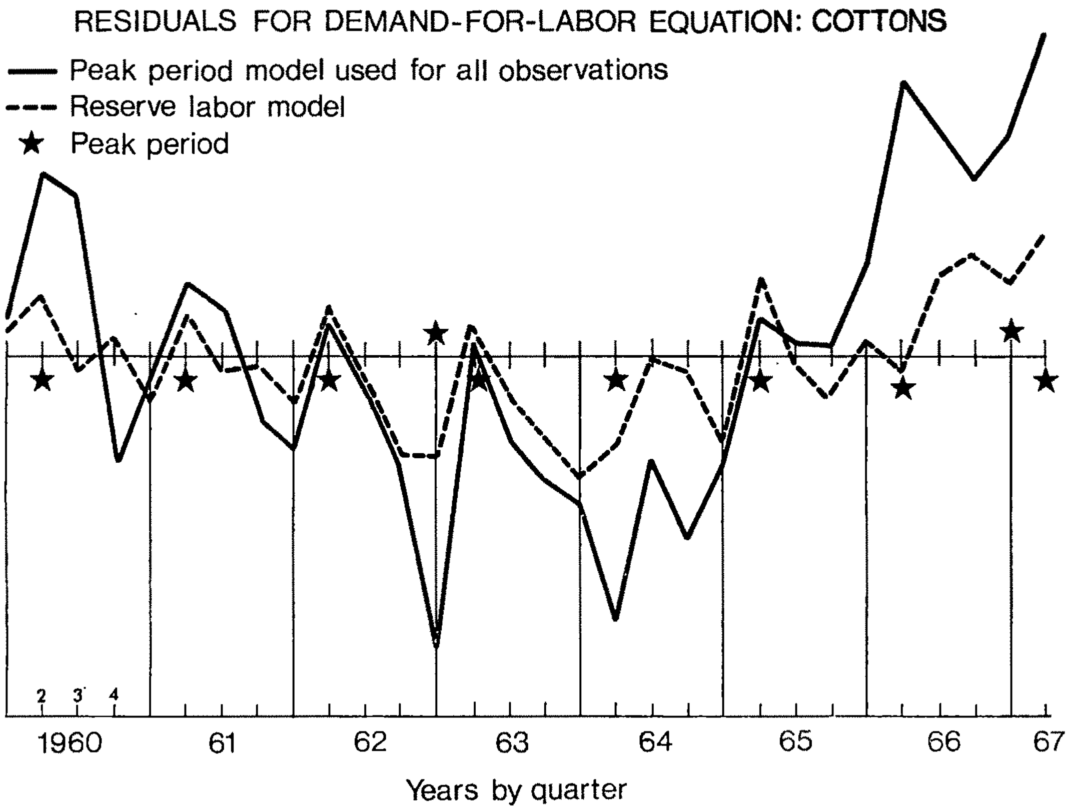


Figure 2

tion that inventories of finished goods are a substitute for inventories of labor.

Finally, I will attempt to assess the value of incorporating the idea of inventories of labor into the system. To do this, I reestimated the system using only the peak-period equations with all observations. The demand-for-labor equation (9′) (without *H*) has been used extensively in empirical work by Frank Brechling, Brechling and Peter O'Brien and others (Ball and St. Cyr, Smyth and Ireland, Robert Soligo). If the reserve labor theory is meaningful, the dichotomized system should be a better predictor of actual labor demand. I therefore computed predicted *log E* for each period from the estimates in the dichotomized system and from the estimates obtained from the peak-period system using all observations. Plots of the residuals are presented in Figures 1 and 2. Residual variances were computed along with Durbin-Watson statistics.[14] They are given in Table 3 below:

TABLE 3—RESIDUAL STATISTICS

	Residual Variance	D. W.
Woolens		
Peak-period system used for all observations	0.001973	1.0472
Reserve labor system with dichotomization	0.000952	1.662
Cottons		
Peak-period system used for all observations	0.002459	0.725
Reserve labor system with dichotomization	0.0005322	1.587

Use of the dichotomized system thus resulted in a reduction in residual variance of about one-half for woolens and about three-fourths for cottons. Moreover, much of the extreme positive serial correlation was eliminated by use of this system.[15] What still persists indicates that more improvement of the system is necessary.

REFERENCES

J. K. Arrow et al., *Studies in the Mathematical Theory of Inventory and Production,* Stanford 1958.

R. J. Ball and E. B. A. St. Cyr, "Short Term Employment Functions in British Manufacturing Industry," *Rev. Econ. Stud.,* July 1966, *33,* 179–207.

Y. Barzel, "The Determination of Daily Hours and Wages," mimeo., Univ. Washington 1969.

G. S. Becker, "Investment in Human Capital: A Theoretical Analysis," *J. Polit. Econ.,* Oct. 1962, Supp., *70,* 9–49.

F. Brechling, "The Relationship between Output and Employment in British Manufacturing Industries," *Rev. Econ. Stud.,* July 1965, *32,* 187–216.

——— and P. O'Brien, "Short-run Employment Functions in Manufacturing Industries: An International Comparison," *Rev. Econ. Statist.,* Aug. 1967, *49,* 277–87.

H. B. Chenery, "Overcapacity and the Acceleration Principle," *Econometrica,* Jan. 1952, *20,* 1–28.

C. C. Holt, "Improving the Labor Market Trade-off between Inflation and Unemployment," *Amer. Econ. Rev. Proc.,* May 1969, *59,* 135–46.

——— et al., *Planning Production, Inventories, and Workforce,* Englewood Cliffs 1960.

——— and M. H. David, "The Concept of Job Vacancies in a Dynamic Theory of the Labor Market," in *The Management and Interpretation of Job Vacancies,* Nat. Bur. Econ. Res. conference report, New York 1966.

---

[14] Comparison of residual variances is not strictly meaningful statistically because of the way predicted *log E* is computed in the dichotomized system. For computation of the D. W. statistic, the calculated residuals from the dichotomized systems were ordered seriatim from 1960 I to 1967 IV.

[15] Note that because of the inclusion of a lagged dependent variable in the demand-for-labor equation used for all observations, the Durbin-Watson statistic is biased toward two. This is not necessarily the case with the dichotomized system, where that equation contains a lagged dependent variable in only about one fourth of the observations.

N. J. Ireland and D. J. Smyth, "The Specification of Short-run Employment Models and Returns to Labor," mimeo. 1967.

M. C. Lovell, "Sales Anticipations, Planned Inventory Investment, and Realizations," in R. Fehher, ed., *Determinants of Investment Behavior*, Nat. Bur. Econ. Res. conference report, New York 1967.

W. Y. Oi, "Labor as a Quasi-fixed Factor of Production, *J. Polit. Econ.*, Dec. 1962, *70*, 538–55.

S. Rosen, "Short-run Employment Variations on Class-I Railroads in the U.S., 1947–63," *Econometrica*, July-Oct. 1968, *36*, 511–29.

———, "On the Interindustry Wage and Hours Structure, *J. Polit. Econ.*, Mar./Apr. 1969, *77*, 249–73.

S. H. Slichter, *The Turnover of Factory Labor*, New York 1919.

D. J. Smyth and N. J. Ireland, "Short-term Employment Functions in Australian Manufacturing," *Rev. Econ. Statist.*, Nov. 1967, *49*, 537–44.

R. Soligo, "The Short-run Relationship between Employment and Output," *Yale Econ. Essays*, spring 1966, *6*, 161–215.

Information and Frictional Unemployment

By Reuben Gronau*

An increasing number of economists discussing the trade off between unemployment and inflation have been trying recently to establish the microeconomic behavior giving rise to the Phillips curve.[1] A great many of these writers (see Armen Alchian, Charles Holt, Dale Mortenson, Edmund Phelps (1968), Melvin Reder) trace this behavior to the scarcity of information. According to this approach the unemployed worker in search of information on job opportunities lowers his wage demands as his search proceeds. Since the length of search is directly related to the level of unemployment, one would expect wage demands (and hence the change in the general wage rate) and unemployment to be inversely related.

This paper tries to reformulate George Stigler's theory of the economics of information (1961, 1962), to examine the factors affecting the job seeker's wage demands. It will be shown that even under conditions of perfect knowledge of future market conditions, maximization of expected income (or expected utility) will lead to a deterioration of the job seeker's real wage demands as the search continues. An increase in unemployment tends to quicken this process. Hence one can derive the Phillips curve without reliance on a lag in information (see Alchian), under-compensation in the setting of wage aspirations (see Holt), or money illusion. We examine the factors making the worker prefer voluntary unemployment to on-the-job search and the factors making him stop his search and drop out of the labor market altogether.[2]

The paper opens with a description of the optimum search policy. The analysis of the factors determining wage flexibility in Section II is followed by a discussion of the optimum combination of direct and indirect costs incurred through search (Section III) and its implication for voluntary unemployment. The concept of "despair" and movement out of the labor force are discussed in Section IV.

I. Optimum Search Policy

A job hunter embarking on search is entering a field of many uncertainties: he does not know what wage offers he will receive, when he will receive them, or how his wage will evolve through time. Even if one ignores the last factor, assuming, as I shall do for the rest of this paper, that on-the-job training and seniority result in an identical rate of increase of all subsequent wages, the job hunter is left with two unknowns. Both the timing and the

---

* Lecturer, the Hebrew University, Jerusalem. I am indebted to Gary S. Becker, Peter A. Diamond, Ruth Klinov-Malul, David Levhari, Jacob Mincer, and Tsvi Ophir and the referee for their useful comments, and to the Maurice Falk Institute for Economic Research for financial support.

[1] For a survey of recent literature, see Edmund Phelps (1969).

[2] While this paper was in the process of publication two other articles discussing the problem of job search and adopting a somewhat similar approach appeared. John McCall is interested mainly in the microeconomic aspect of the problem while Mortensen carries the analysis forward to derive a formal relationship interpreted as the Phillips curve. Both assume that, market conditions remaining unchanged, the job seeker's demands stay constant throughout the search, thus overlooking one of the more important features of the search process.

level of wage offers are subject to random distributions.

Given a job hunter with a finite time horizon, the time span can be broken into N discrete intervals, each small enough for the probability of more than one wage offer arriving during it to approach zero. The job hunter places a probability of p_n on the arrival of a wage offer during time interval n, the wage offer W_n being subject to a random cumulative distribution $F_n(W)$.[3] The variable W_n consists of money earnings (expressed in real terms) as well as of the psychic earnings the worker expects to derive from the job.

Search is a sequential process evolving through time. In each period the job hunter decides on his asking wage w_n. He stops his search if he receives an offer exceeding (or equal to) the asking wage $W_n \geq w_n$; otherwise he continues his search.[4]

The asking wage in period n is not independent of the asking wages the job seeker intends to charge in the future. Clearly, the length of search and the wage offer accepted depend on the asking wage and how it changes while the search continues. The crucial problem facing the job hunter is the formulation of an optimum strategy determining the asking wage at each point of time.

The asking wage will never fall short of any alternative option open to the job hunter. The optimum asking wage in period n should be set in such a way that any wage offer accepted during the period assures the worker of an income stream not inferior to the expected income stream he might have got had he continued his search. Specifically, let R_n be the present value of the income stream generated by a one dollar wage offer accepted in period n; then, given a job hunter who has no time preference, the utility derived from $R_n w_n$ should at least equal the expected utility yielded by continuing the search. Assuming that the job hunter is insensitive to risk, the necessary condition for the optimum asking wage requires the present value of the income stream generated by the asking wage, $R_n w_n$, to equal or exceed the income stream expected if the search is pursued beyond time n.[5] Let I_n denote the expected value of continuing the search beyond point n (i.e. the expected income stream), then the necessary condition for the optimum asking wage is $R_n w_n \geq I_n$ or $w_n \geq I_n / R_n$.

The alternative options open to a job hunter depend on his alternative earnings in the market and in the non-market sector and on how long each offer is open. The duration of offers determines, at least partly, the number of offers there are to choose from at each point of time. Since the job seeker will never accept any offer that falls short of the highest unexpired wage offer received in the past, he will tend to be more choosy, i.e., his asking wage will tend to be higher, the longer offers are open.[6]

The duration of offers determines the rate of obsolescence of the investment in information. It may vary from infinity down to zero. At the one extreme, the job seeker can return to any offer received in the past whenever it suits him, with no danger of the offer having lapsed in the meantime.[7] At the other extreme, the job seeker is faced with an instantaneous decision—an offer not accepted is an offer lost.

---

[3] I assume that p_n is independent of $F_n(W)$, and that both are independent of the corresponding distributions in other periods. The assumption of a discrete time interval is not essential to the analysis and is adopted only for simplification. For an alternative approach, see Gustav Elfing.

[4] Note the difference between this model and Stigler's (1962). Stigler tacitly conceives of a worker who decides beforehand on the number of searches he will undertake, and who sticks to his decision regardless of the outcome of the search. For further analysis of the differences between Stigler's model and the sequential model, see P. Nelson's unpublished paper (some of which is included in his 1970 article).

[5] Both the assumptions of no time preference and insensitivity to risk can be relaxed at the expense of complicating the model.

[6] I assume that the fixed costs associated with search (both pecuniary and psychological) are sufficiently high to rule out the possibility of perpetual job changing, the worker moving from one job to a better paid one as it comes along.

This paper adopts the last assumption, assuming in effect that the search process has no memory. Every offer expires as soon as it is made; the worker has no alternative job offers to fall back on, and his asking wage in period n is, therefore, unaffected by past offers.

An employed job seeker will never accept any wage offer that falls below his current wages (where the wages include the money equivalent of the worker's psychic income). Similarly, an unemployed person will never consent to any wage offer that does not exceed the value of his marginal product outside the market sector (where this value includes the money equivalent of the utility derived from leisure) plus any unemployment insurance he may receive.

Let W^n denote the job seeker's alternative earnings in the market and non-market sector in period n ($W^n \geq 0$), then the asking wage w_n will never fall short of W^n. Since the worker has no incentive to offer his services at a price lower than the minimum value of the wage offers distribution $\underline{W}_n$ [where $\underline{W}_n = \text{Min } W$ satisfying $dF_n(W) > 0$], W^n will serve as the lower boundary of the asking wage, as long as it exceeds $\underline{W}_n$. When it does not, it will be replaced by $\underline{W}_n$.[8] Hence, the necessary and sufficient condition for an optimum asking wage is

$$w_n = \text{Max } (I_n/R_n, W^n, \underline{W}_n) \tag{1}$$

Given the above assumptions, the optimum asking wage depends on W^n, $\underline{W}_n$, R_n, and I_n. The current wage W^n and the minimum wage offered $\underline{W}_n$ are parameters that can be assumed, as a first approximation, to be given exogenously.[9] In equation (2), the variable R_n is the present value of the income stream generated by a one dollar wage offer accepted in period n. Its value depends on the number of periods the worker intends to stay in his job (N−n), on the real rate of interest (r), and on the (real) rate of wage advances (β). In equation (2), the income stream is discounted to the initial period 0, and the worker is assumed to start working as soon as he accepts the offer.

The probability that a job offer will be received in period n is p_n. The probability that an offer received will be accepted is G_n where

$$G_n = \text{Prob } (W_n \geq w_n) = 1 - F_n(w_n) \tag{3}$$

The probability that an offer will be received and accepted in period n therefore equals p_nG_n. Let E_n denote the average ac-

$$R_n \begin{cases} = \dfrac{1}{(1+r)^{\text{N}}} \dfrac{(1+r)^{\text{N}-\text{n}+1} - (1+\beta)^{\text{N}-\text{n}+1}}{r-\beta} & \text{when } r \neq \beta \\ = \dfrac{1}{(1+r)^{\text{n}}} (\text{N} - \text{n} + 1) & \text{when } r = \beta \end{cases} \tag{2}$$

---

[7] This kind of assumption may have been at the back of Stigler's analysis when he discusses the returns of the search. Still, even in this case, the sequential model should be preferable to Stigler's static analysis.

[8] The notation used in the text implies that W^n does not necessarily belong to the distribution $F(W)$, i.e., the job seeker's current earnings may be outside the range of the job offers distribution (e.g., the job seeker is unemployed).

[9] As a first approximation, I ignore the interaction between the behavior of workers looking for jobs and that of employers looking for workers.

ceptable wage in period n. The variable E_n is the expected value of W_n given that it exceeds w_n.

$$\begin{aligned} E_n &= E(W_n/W_n \geq w_n) \\ &= \frac{1}{G_n}\int_{w_n}^{\infty} W dF_n(W) \end{aligned} \tag{4}$$

A job seeker entering period n places a probability of p_nG_n on finding within the period, a job that will yield him on the average an expected income of R_nE_n.

There is a probability of $1-p_nG_n$ that a job will not be found in period n. Search is a time-consuming process and thus may very well result in a decline of the job hunter's current earnings. Absenteeism or a decline in the time inputs in the home sector may reduce earnings to αW^n $(0 \leq \alpha \leq 1)$ as long as the search continues. Thus, if no job is found in period n, the job seeker will enjoy an income of $\alpha W^n/(1+r)^n = [R_n-(1+\beta)R_{n+1}]\ \alpha W^n$ during the period, and can look forward to an expected income of I_n in subsequent periods.

The variable I_{n-1} is the income stream the worker expects as he enters period n. Ignoring the direct costs of search, it equals

$$\begin{aligned} I_{n-1} = p_nG_nR_nE_n &+ (1-p_nG_n) \\ &\cdot \{[R_n - (1+\beta)R_{n+1}]\alpha W^n + I_n\} \\ &\qquad n = 1, \ldots, N, \end{aligned} \tag{5}$$

where $I_N = 0$.

Assuming the job seeker does not stop his search prematurely, one can insert the values of $I_n, I_{n+1}, \ldots, I_N$ in equation (5) to obtain

$$\begin{aligned} I_{n-1} = R_n\alpha W^n &+ R_n(E_n - \alpha W^n)p_nG_n \\ &+ \sum_{k=n+1}^{N} R_k[E_k - (1+\beta)^{k-n}\alpha W^n] \\ &\cdot p_kG_k \prod_{i=n}^{k-1}(1 - p_iG_i) \\ &\qquad n = 1, \ldots, N, \end{aligned} \tag{6}$$

since $W^{n+1} = (1+\beta)W^n$ and $I_N = R_{N+1} = 0$. The job seeker entering period n expects with certainty to enjoy an income of $R_n\alpha W^n$ for the rest of his time horizon, and there exists a chance, measured by the probability

$$p_kG_k \prod_{i=n}^{k-1}(1 - p_iG_i),$$

that a new job will be obtained in period k, increasing his wealth, on the average, by $R_k[E_k-(1+\beta)^{k-n}\alpha W^n]$. The expected value of search is a function of the future search strategy. Note, however, that the weight of future decisions declines the further they are removed from the present. Decisions made in the distant future may, therefore, have only a negligible effect on the job seeker's behavior in the present.

II. The Factors Determining Wage Flexibility

How does the asking wage w_n change through time? Changes in the asking wage may be attributed to two sources: (a) changes that occur as part of the optimum search strategy, and (b) changes that result from a modification of the optimum strategy. A modification of strategy may take place because of unexpected changes in the labor market or because hitherto unknown information has become available in the course of the search.[10] For simplicity, it will be assumed that there exists perfect foresight on future conditions in the labor market [p_n, $F_n(W)$, etc.]. The worker knows the general shape of the wage offer distribution $F_n(W)$, but does not know which employer has a vacancy or the wage offer associated with each vacancy.

To obtain the optimum path of the asking wage, let us assume that I_n/R_n exceeds both $\underline{W}_n$ and W^n. By equation (5)

---

[10] This is essentially the case that Alchian has in mind. For a further discussion of this case, see Joseph A. Yahav.

$$
\begin{aligned}
w_{n-1} &= I_{n-1}/R_{n-1} \\
&= \frac{R_n}{R_{n-1}}\Big[w_n + p_nG_n(E_n - w_n) \\
&\qquad + (1 - p_nG_n)\frac{\alpha W^n}{(1+r)^nR_n}\Big] \qquad (7)\\
&\qquad\qquad n = 1, \ldots, N-1
\end{aligned}
$$

The asking wage in period n−1 is a function of current earnings, the asking wage in the next period, and the average acceptable wage in the next period.

The search must terminate within N periods ($I_N = 0$). Hence, the job seeker cannot be choosy in the last period, his asking wage being equal to the minimum wage offer or to the alternative earnings, whichever is higher:

$$
w_N = W_N^*, \tag{8}
$$

where $W_n^* = \text{Max}\ (W^n, \underline{W}_n)$. Assuming $I_n/R_n > W_n^*$ for all n=1, . . . , N−1, the asking wage in period N−1 equals, by equation (5)

$$
w_{N-1} = \frac{R_N}{R_{N-1}}\left[p_NG_NE_N + (1 - p_NG_N)\alpha W^N\right] \tag{9}
$$

Combining this initial condition with equation (7), one may obtain a recursive solution for all values of w_n (n=1, . . . , N−2).

The exact nature of the asking wage path depends, of course, on the specific values of the parameters in equation (7). Still, one can trace some general features of the job seeker's behavior. The asking wage is determined by the ratio of the expected income from search (I_n) to the present value of the income stream generated by a one dollar wage offer accepted in period n (R_n). Both these terms decline as the search proceeds. However, assuming that the rate of job offer arrivals (p) and the wage offer distribution [$F(W)$] remain constant throughout the search, and that the job seeker enjoys no earning in the present ($W^n = 0$), it can be shown that I_n declines at a faster rate than R_n, resulting in a fall of the asking wage over time.[11]

Other parameters remaining constant, the factor responsible for the decline in the asking wage is the finite time horizon. The job seeker's time horizon depends on the time he expects to stay on his next job. A job seeker embarking on search may expect (as I have assumed) to stay on his next job until a specified date (i.e., to a given period N), or, alternatively, for a specified length of time (i.e., for a given number of periods N−n). Given the first assumption, any prolongation of search will cut directly into the time spent on the next job. By the second assumption, the length of search does not affect the duration of employment on the next job, but, given the finite life span, will necessarily affect the time spent on some subsequent jobs. Either way, any prolongation of the search reduces the gains to be expected from the search.[12] The average period of employment in the United States is about 2.7 years, and the average period of unemployment is about one month (see Holt, p. 137), so that this reduction may be quite significant.

A major determinant of the asking wage path is the distribution of wage offers $F(W)$. It determines the asking wage threshold $w_N = \underline{W}$, and, given the asking wage in period n (w_n), the probability of acceptance G_n and the average acceptable wage E_n, two factors that are crucial in setting the asking wage in the preceding period (w_{n-1}).

Changing the scale of the distribution

---

[11] See Appendix.

[12] Mortensen explicitly assumes an infinite time horizon. McCall in essence adopts the assumption that the job seeker intends to stay on his job for a specified length of time but overlooks the effect of prolongation of search on subsequent jobs. Thus, both ignore the cost of elapsed time involved in search, and conclude that the asking wage remains constant throughout the search.

(i.e., multiplying the values of W and W^n by the same constant) does not affect the job seeker's behavior (i.e., w_n is multiplied by the same constant). However, given the center of location of the distribution (e.g., its mean), the expected gains from search increase with dispersion of the distribution (e.g., its standard deviation). Alternatively, given the dispersion of the distribution, the foregone earnings associated with continuing the search increase with the location of the distribution. Furthermore, the job seeker's optimum strategy may be influenced also by higher moments of the distribution, most notably, the symmetry of the distribution—the gains from search being greater the more skewed the distribution is to the left. Consequently, one would expect the deceleration of wage demands to increase with the positive skewness of the distribution and to decrease with its relative variation (i.e., its coefficient of variation).

The job seeker's chances of obtaining a new job depend, to a large extent, on the job offer stream. The job seeker's demands increase, therefore, with the rate of arrival of job offers (p).

The importance of the time of entry into the new job increases with the rate of interest (r) and the rate of wage advances (β). The higher the personal rate of discount, the lower the present value of any possible gains to be reaped by search, and, hence, the greater the urgency of getting a job fast. Similarly, the greater the remuneration for seniority and on-the-job training (i.e., the higher β), the greater the foregone earnings associated with continuing the search, and, hence, the greater the worker's tendency to accept lower wage offers at any point of time.

The incorporation of current earnings in the model cannot change the major conclusions reached above. Current earnings (W^n) increase the job hunter's boldness. The higher these earnings and the higher the percentage of earnings retained throughout the search (α), the smaller the loss of income associated with the search. Furthermore, when W^n exceeds $\underline{W}$ current earnings set the floor to any wage acceptable. Thus the higher the job hunter's current income, the higher his asking wage. An employed person can, of course, be more choosy than an unemployed one.

The search policy and market conditions determine the average length of search S_n.

$$S_n = p_nG_n + (1 - p_nG_n)(1 + S_{n+1}) = p_nG_n + \sum_{k=n}^{N} (k - n - 1) p_kG_k \prod_{i=n}^{k-1} (1 - p_iG_i) \tag{10}$$

The bolder the policy adopted by the job seeker, the longer he can expect the search to last. Given that the worker's demands fall as the search continues, so does the expected average length of search. Other things being equal, the length of search and the time horizon are inversely related.

The job seeker charges a higher wage the higher his current earnings, the higher the percentage of earnings retained throughout the search, and the greater the relative variation of the wage offer distribution. All these factors prolong the search. An increase in the rate of interest, on the other hand, moderates the job seeker's demands and curtails the duration of search. Finally, the higher the rate of unemployment, the lower the asking wage. It seems that the decline in wage demands does not offset completely the slower rate of job offers, however, and as a result a decline in p becomes a major factor in the prolongation of the search.[13]

How does an increase in the market rate of unemployment affect the wage demands of labor? An increase in the rate of unemployment, or a fall in the rate of vacancies, reduce the job offer stream and, hence, accelerate the fall of the job seeker's demands. This tendency may be accentuated by several additional changes: (a) the increase in unemployment may be associated with a shorter employment period and, hence, a shorter time horizon; (b) the rate of interest facing the job seeker is lower the better he is equipped with assets (and, in particular, liquid assets). Prolongation of search may result in a deterioration of the unemployed worker's asset position, and in a subsequent increase of his personal discount rate; (c) when the job seeker is unemployed, W^n denotes his marginal productivity in the non-market sector and the money equivalent of his marginal utility of leisure. It has been suggested that a prolonged period of unemployment leads to a decrease of the marginal utility of leisure (see Hirschel Kasper) and, thus, to a decline of W^n; finally (d) the recession may be accompanied by a decline of the center of location of the wage offers distribution and by an increase in its positive skewness, "good" offers becoming more difficult to come by.

---

[13] For a more detailed analysis see Mortenson.

Thus, periods of increased unemployment are accompanied by a decline in the real wage demands of labor whether the recession is anticipated or not. The fall in the asking wage will be more abrupt when worsened employment conditions set in unexpectedly. However, one would also expect a gradual erosion of labor demands when the recession is expected, even if workers suffer from no money illusion, have perfect knowledge of the new market conditions, and there is no time lag between the change in prices and the change in wages.

III. Voluntary Unemployment

The discussion so far has given only a partial description of the factors bearing on the job hunter's optimum strategy. It depicts the factors affecting the length of search but does not explain the factors determining the intensity of search, i.e., the amount of resources devoted to search at each period.

The stream of job offers is a function of market conditions such as the market size and the state of employment, as well as of the amount of time and money resources applied to the search. Given the state of the market, the worker can accelerate or decelerate the rate of arrival of job offers by increasing or decreasing his direct or indirect costs. Direct costs of search include payments for employment agencies and advertisements, while the indirect costs consist mainly of foregone earnings.[14] If more time is spent on search, less time can be devoted to other activities. Since search seems to be more efficient when it takes place during regular working hours, an increase in time spent in search may very well come at the expense of time spent on the job, so that current earnings decline. The exact nature of the decline in current earnings depends on the amount of time diverted to the search, as well as on some specific factors, such as the employer's leniency towards absenteeism, the employee's opportunity to search during vacation, etc. An optimum search policy is expected to yield answers not only to the question how long to search, but also to questions about the optimum combination of inputs and the amount of resources that should be invested in the process.

Formally, the rate of arrival of job offers (p) is not a predetermined parameter, but an endogenous variable. It is a function of past and present investment of time and market inputs.[15] Let C denote direct costs,

---

[14] Payments to employment agencies may be in the form of a fixed share of earnings (say, one-half of the first month's salary). One can take account of this by an appropriate modification of R_n.

[15] I ignore any effect the investment in search may have on the shape of the wage offer distribution $F(W)$. An employment agency may serve as a screening device, passing only the best offers available. In this case, $F(W)$ becomes the distribution of the maximum values.

$$
\begin{aligned}
I_{\mathrm{n}-1} = {} & R_{\mathrm{n}}W^{\mathrm{n}} + p_{\mathrm{n}}G_{\mathrm{n}}R_{\mathrm{n}}(E_{\mathrm{n}} - W^{\mathrm{n}}) + \sum_{\mathrm{k}=\mathrm{n}+1}^{\mathrm{N}} R_{\mathrm{k}}(E_{\mathrm{k}} - W^{\mathrm{k}})p_{\mathrm{k}}G_{\mathrm{k}} \prod_{i=\mathrm{n}}^{\mathrm{k}-1} (1 - p_iG_i) \\
& - \left[\frac{C_{\mathrm{n}}}{(1+r)^{\mathrm{n}}} + \sum_{\mathrm{k}=\mathrm{n}}^{\mathrm{N}-1} \frac{1}{(1+r)^{\mathrm{n}}}\left(\frac{C_{\mathrm{k}+1}}{1+r} + (1-\alpha_{\mathrm{k}})W^{\mathrm{k}}\right) \prod_{i=\mathrm{n}}^{\mathrm{k}} (1 - p_iG_i) \right. \\
& \left. + \frac{(1-\alpha_{\mathrm{N}})W^{\mathrm{N}}}{(1+r)^{\mathrm{N}}} \prod_{i=\mathrm{n}}^{\mathrm{N}} (1 - p_iG_i)\right] \qquad \mathrm{n} = 1, \ldots, \mathrm{N}
\end{aligned} \tag{11}
$$

and $(1-\alpha)$ be the percentage of current income lost, then[16]

$$p = g[(1-\alpha), C]$$

is the job offers production function describing the various combinations of time and money resources required to produce a given stream of job offers. The job hunter has to determine not only his asking wage at each period (w_{n}), but also the amount of time and market input ($1-\alpha_{\mathrm{n}}$ and C_{n}, respectively) applied to the search. Put differently, he must choose the optimum set of policy variables $w=(w_1 \ldots w_{\mathrm{N}})$, $\alpha=(\alpha_1 \ldots \alpha_{\mathrm{N}})$, and $C=(C_0, C_1, \ldots C_{\mathrm{N}})$ so as to maximize $I_{\mathrm{n}-1}$ as shown in equation (11), subject to $p=g[(1-\alpha), C]$. The first part of equation (11) describes the expected gains from search and the second part (the terms in brackets) describes the costs.

The job seeker will combine the time and market inputs in such a way that their marginal rate of substitution in the production of job offers equals their relative shadow prices. If one assumes that the investment in search in period n does not affect the stream of job offers in subsequent periods (i.e., $\partial P_{\mathrm{k}}/\partial C_{\mathrm{n}} = \partial p_{\mathrm{k}}/\partial(1-\alpha_{\mathrm{n}}) = 0$ for any $\mathrm{k} > \mathrm{n}$), the optimum combination of direct and indirect costs is obtained when

$$
\begin{aligned}
g_{(1-\alpha)}/g_C &= \frac{\partial I_{\mathrm{n}}/\partial(1-\alpha)}{\partial I_{\mathrm{n}}/\partial C} \\
&= (1 - p_{\mathrm{n}}G_{\mathrm{n}})W^{\mathrm{n}}
\end{aligned} \tag{12}
$$

where $g_{(1-\alpha)}$, and g_C denote the marginal products of time and market input, respectively, in the production of job offers. Other things being equal, the tendency to substitute direct costs for indirect costs is stronger the higher the job seeker's current income (W^{n}), and the slighter the chances of accepting a new job (i.e., the smaller p and G). Thus, one would expect skilled workers to spend more on employment agencies and advertising than unskilled workers. Moreover, one would expect the job seeker to get more personally involved in the process of search (i.e., spend more time on search) as the process evolves and his asking wage declines (i.e., G_{n} increases).

Vacations being rare and employers' leniency towards absenteeism being limited, there may exist a maximum to the number of work hours the employee may miss, i.e., there may exist some lower bound to α (say $\underline{\alpha} > 0$). If the optimum α_{n} falls short of $\underline{\alpha}$, the job seeker will prefer voluntary unemployment to on-the-job search (see Alchian).[17] Thus, given equation (12), one would expect voluntary unemployment to be more prevalent the lower the worker's earnings, and the more time has already been spent in search (the greater G_n). Finally, worker's tendency to quit their job in order to concentrate on search increases the greater the job offer stream p. Voluntary unemployment should therefore be more common in periods of rising aggregate demand, in particular if the rise in demand and profits is accompanied by a fall in employers' tolerance of absenteeism (i.e., a rise in $\underline{\alpha}$).

---

[16] The variables p, $(1-\alpha)$, and C can be interpreted as vectors. The rate of job offer arrivals in period n is a function of past and present investments in search $p_{\mathrm{n}} = g(C_0, C_1, \ldots, C_{\mathrm{n}}, \alpha_1, \ldots, \alpha_{\mathrm{n}})$.

[17] Both McCall and Mortensen ignore on-the-job search. However, this kind of search is far from negligible. Robert L. Stein found that in 1955, about 55 percent of all job changers did not suffer any unemployment.

IV. Movement Out of The Labor Force

One of the assumptions made in Section II is that the job seeker does not stop his search prematurely. This assumption, ruling out despair, may prove too restrictive and, hence, will be relaxed in this section.

A job hunter stops his search altogether when he finds that the present value of his current earnings exceeds the expected value of the search. The present value of the earnings of a worker who stops his search in period n−1 is $R_n W^n$. Hence, the necessary and sufficient condition for despair is

$$(13) \quad R_n W^n > I_{n-1} \text{ or } W^n > I_{n-1}/R_n$$

When the rate of wage advances equals zero ($\beta=0$), $W^n=W^0$ for every n, and since $R_{n-1}>R_n$, a necessary condition for the termination of the search is

$$W^0 > I_{n-1}/R_{n-1},$$

that is, the job seeker does not stop his endeavors as long as his asking wage exceeds W^*_{n-1}.[18]

Let the rate of job offer arrivals (p) and the wage offer distribution stay constant through time (i.e., $W_n^*=W^*$ for every n). A job seeker hesitating whether to stop his search in period n−1 or whether to continue it for one additional period, will stop if

$$(14) \quad \begin{aligned} R_n W^0 > I_{n-1} &= pG^* R_n E^* \\ &+ (1-pG^*)[(R_n - R_{n+1})\alpha W^0 + R_{n+1} W^0], \end{aligned}$$

where

$$G^* = 1 - F(W^*) \text{ and } E^* = E(W/W>W^*)$$

Ignoring direct costs, he will drop out of the search when the expected future gains of the search cannot compensate him for the income lost in the present.

$$(15) \quad \begin{aligned} &(1-\alpha)W^0 > pG^* \\ &\left[E^* - \alpha W^0) + \frac{R_{n+1}}{R_n - R_{n+1}}(E^* - W^0)\right], \end{aligned}$$

$\alpha<1$ being a necessary condition for despair.

Put differently, despair sets in as soon as

$$(16) \quad \frac{R_{n+1}}{R_n} < 1 - \frac{pG^*}{1 - pG^*} \frac{E^* - W^0}{(1-\alpha)W^0}$$

where $p < 1$, $\alpha < 1$.[19]

The job seeker is more susceptible to despair the higher his current earnings W^0, the greater the loss of income due to search $(1-\alpha)$, the smaller the wage dispersion (i.e., the smaller E^*-W^0), the higher the

---

[18] The wage charged in period N equals $W^*_N=\text{Max}(W^N, W^N)$ and the asking wage in period N−1 should have been

$$I_{N-1}/R_{N-1} = \frac{R_N}{R_{N-1}}[p_n G_N E_N + (1 - p_N G_N)\alpha W^N]$$

However, given values of r and β that are sufficiently large (i.e., R_N/R_{N-1} being sufficiently small), and values of α, p_N, and G_N that are sufficiently small, there is no certainty that I_{N-1}/R_{N-1} exceeds W^*_{N-1}, and if it does not, the asking wage in period N−1 will equal the minimum acceptable wage W^*_{N-1}. Similarly, one can conceive of a case where the asking wage hits its floor in period n, and stays at the minimum level for the remaining (N−n) periods.

[19] The right-hand side of the inequality is independent of n. The validity of the inequality depends, therefore, on the left-hand side. R_{n+1}/R_n is a decreasing function of n. Thus, if the inequality holds for n, it will hold for every k>n. For example, when $r=0$ the worker drops out of the market as soon as

$$n > (N+1) - \frac{1 - pG^*}{pG^*} \cdot \frac{(1-\alpha)W^0}{E - W^0}$$

rate of interest (r), and the slower the rate of job offer arrivals (p).

When current alternative earnings fall short of the minimum wage offer (i.e., $W^0 < \underline{W}$), the wage charged is the minimum wage offered; $W^* = \underline{W}$ and hence $G^* = 1$ and $E^* = \mu$, where μ is the mean of the wage offer distribution. Rewriting equation (16), the search is stopped when

$$\frac{R_{n+1}}{R_n} < 1 - \frac{p}{1-p} \cdot \frac{\mu - W^0}{(1-\alpha)W^0}, \tag{17}$$

$p < 1$ being a necessary condition for despair.

An economic recession results in a dwindling stream of job offers (a decline in p), a downward shift in the job offers distribution (a decline in μ and E^*), and a possible increase in the personal rate of interest (r). All these factors have a discouraging effect on search. Employed persons will tend, therefore, to stick to their jobs, and the unemployed will be more easily tempted to drop out of the labor force altogether. The chances of finding a job and the difference between market earnings (μ) and the value of marginal product outside the market sector (W^0) are smaller for women than for men to start with. Women are the first to bear the brunt of unemployment, thus giving rise to the well-known "discouraged worker hypothesis."

On the other hand, when a husband loses his job, his wife may decide to enter the labor market. Having the husband at home reduces the value of the wife's marginal product, and the loss of this product suffered because of the search. The decline in W^0 and $(1-\alpha)$ increases the profitability of search, giving rise to the "added worker hypothesis."

V. Conclusion

In the preceding four sections an attempt was made to reformulate the theory of search and to apply it to some problems associated with frictional unemployment. It was shown that under constant market conditions, expected income maximization leads to a declining asking wage. Unemployment tends to accelerate the fall of job seekers' demands, and hence may lead to a Phillips curve-like relationship between wage change and unemployment. Costs of search minimization may lead the job seeker to substitute his own time for market inputs. At the extreme, the job seeker may prefer voluntary unemployment to on-the-job search to speed up the stream of job offers confronting him. A formalization of the concept of "despair" led to the analysis of the job seeker's incentives to join or quit the search (and labor force).

The search model can easily be adapted to describe employers' behavior in trying to fill vacancies. A more ambitious task is the application of the theory to the analysis of the expected period of employment, and hence, quit rates and discharge rates. Finally, one would like to carry the analysis from the individual level to the market level to analyze the determinants of the wage offer and the asking wage distributions as well as the actual distribution of earnings. These tasks still lie ahead.

APPENDIX

By equation (5)

$$\begin{aligned} I_{n-1} &= p_n G_n R_n E_n + (1 - p_n G_n) R_n \\ &\quad \cdot \left[\left(1 - (1+\beta) \frac{R_{n+1}}{R_n} \right) \alpha W^n + w_n \right] \end{aligned} \tag{1'}$$

Assuming that $W^n = 0$ and that p_n and $F_n(W)$ remain constant over time

$$I_{n-1} = R_n [p G_n E_n + (1 - p G_n) w_n] \qquad n = 1, \ldots, N \tag{2'}$$

It must be shown that $w_{n-1} > w_n$, i.e., $I_{n-1}/R_{n-1} > I_n/R_n$. Employing a proof by induction: $I_N = 0$, and $I_{N-1} = p_N G_N R_N E_N > 0$; hence, the statement is true for $k = N$.

Let the statement be true for $k = n+1 (w_n > w_{n+1})$, and show it is true for $k = n$. Let us assume that in period n, the job seeker charges an asking wage of $w_n' = w_{n+1}$ instead of the optimum wage w_n. The expected income in period $n-1$ is

$$
\begin{aligned}
I'_{n-1} &= R_n[pG'_n E'_n + (1 - pG'_n)w_n] \\
&= R_n[pG_{n+1}E_{n+1} + (1 - pG_{n+1})w_n] \qquad (3')
\end{aligned}
$$

By the induction assumption

$$
\begin{aligned}
I'_{n-1} &> R_n[pG_{n+1}E_{n+1} + (1 - pG_{n+1})w_{n+1}] \\
&= R_n[I_{n+1}/R_{n+1}] \qquad (4')
\end{aligned}
$$

and hence

$$I'_{n-1}/I_n > R_n/R_{n+1} \qquad (5')$$

But I'_{n-1} originates in a sub-optimum strategy and hence $I_{n-1} > I'_{n-1}$. By equation (2)

$$
\frac{R_n}{R_{n+1}} = \begin{cases} \dfrac{(1+r)^{N-n+1} - (1+\beta)^{N-n+1}}{(1+r)^{N-n} - (1+\beta)^{N-n}} & \text{when } r \neq \beta \\[2ex] \dfrac{(1+r)^{n+1}(N-n+1)}{(1+r)^n(N-n)} & \text{when } r = \beta \end{cases} \qquad (6')
$$

Hence $R_n/R_{n+1} > R_{n-1}/R_n$, and

$$
\begin{aligned}
I_{n-1}/I_n &> R_{n-1}/R_n \Rightarrow \\
I_{n-1}/R_{n-1} &> I_n/R_n \qquad \text{Q.E.D.} \qquad (7')
\end{aligned}
$$

The asking wage w_n is a decreasing function of n for any log-convex function R_n, regardless of the shape of $F(W)$.

By equation (2′)

$$w_{n-1} = \frac{R_n}{R_{n-1}}[w_n + pG_n(E_n - w_n)] \qquad (8')$$

Other things being equal, the asking wage increases with the rate of job offer arrival (p), the difference between the average acceptable wage and the asking wage ($E_n - w_n$) and R_n/R_{n-1}. The second of these terms is directly related to the wage offers distribution's relative variation and to its leftward skewness. The last term (R_n/R_{n-1}) is inversely related to both the rate of interest (r) and the rate of wage advances (β).

When $W^n \neq 0$ one has to correct (8′) to obtain

$$
\begin{aligned}
w_{n-1} = \frac{R_n}{R_{n-1}}\Big[&w_n + pG_n(E_n - w_n) \\
&+ (1 - pG_n)\frac{\alpha W^n}{(1-r)^n R_n}\Big] \qquad (9')
\end{aligned}
$$

The effect current earnings (W^n) have on the asking wage increases with α and decreases with p, w_n and R_n (i.e., with $(N-n)$). When R_n is sufficiently large, or α sufficiently small, the optimum strategy in the case of $W^n \neq 0$ approaches the policy adopted with $W^n = 0$.

Finally, when p_n and $F_n(W)$ do not remain constant over time, one can always choose p_n small enough so that $w_{n-1} < w_n$. Similarly, one can assume that the distribution $F_n(W)$ lies to the left of $F_{n+1}(W)$ making the difference ($E_n - w_n$) sufficiently small to yield the same result.

REFERENCES

A. A. Alchian, "Information Costs, Pricing, and Resource Unemployment," *Western Econ. J.*, June 1969, *7*, 109–28.

G. Elfving, "A Persistency Problem Connected with a Point Process," *J. Appl. Probability*, Apr. 1967, *4*, 77–89.

C. H. Holt, "Improving the Labor Market Trade-off between Inflation and Unemployment," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 135–46.

H. Kasper, "The Asking Price of Labor and the Duration of Unemployment," *Rev. Econ. Statist.*, May 1967, *49*, 1965–72.

J. J. McCall, "Economics of Information and Job Search," *Quart. J. Econ.*, Feb. 1970, *84*, 113–26.

D. T. Mortensen, "Job Search, the Duration of Unemployment, and the Phillips Curve," *Amer. Econ. Rev.*, Dec. 1970, *60*, 847–62.

P. Nelson, "Information and Consumer Behavior," unpublished.

———, "Information and Consumer Behavior," *J. Polit. Econ.*, Mar./Apr. 1970, *78,* 311–29.

E. S. Phelps, "Money-Wage Dynamics and Labor-Market Equilibrium," *J. Polit. Econ.*, July-Aug. 1968, part II, *76,* 678–711.

———, "The New Microeconomics in Inflation and Employment Theory," *Amer. Econ. Rev. Proc.*, May 1969, *59,* 147–60.

M. W. Reder, "The Theory of Frictional Unemployment," *Economica,* Feb. 1969, *36,* 1–28.

R. L. Stein, "Unemployment and Job Mobility," *Mon. Lab. Rev.*, Apr. 1960, *83,* 350–58.

G. J. Stigler, "The Economics of Information," *J. Polit. Econ.*, June 1961, *69,* 312–25.

———, "Information in the Labor Market," *J. Polit. Econ.*, Oct. 1962, Supp., *70,* 94–105.

J. A. Yahav, "On Optimal Stopping," *Ann. Math. Statist.*, Feb. 1966, *37,* 30–35.

The Efficient Allocation of Subsidies to College Students

By STEPHEN A. HOENACK*

A current problem of public policy is to increase enrollments from disadvantaged groups in high quality colleges and universities. Even when students from disadvantaged groups are academically qualified, additional financial inducements are required to increase their enrollments, leaving less money to subsidize other groups. This paper attempts to illustrate the structure of tuition charges that will simultaneously achieve the objectives of total enrollment and the correct distribution of students from various income groups, given a fixed total subsidy available to the university.

The paper examines alternative ways of allocating subsidies among college students from different family income backgrounds. Since the University of California (the University) charges fees which are substantially below costs, all of its students receive an implicit subsidy.[1] The aggregate subsidy could be allocated in alternative ways by giving smaller subsidies to some groups of students and larger subsidies to other groups through differential tuition charges. Whether a particular alternative subsidy allocation increases or decreases the total level of enrollment depends on the nature of each group's enrollment demand curve.[2] The data for 1967 show that California high school seniors from different parental income groups were unequally represented at the University. Attendance was greater in the higher income groups. Estimated demand functions for attendance at the University show that a subsidy reallocation aimed toward a more equal representation from the income groups would increase total enrollment somewhat if the reallocation was small. However, beyond some point, a reallocation would bring about a smaller total enrollment, because successive movements towards equality become increasingly more costly in terms of reductions in the level of total enrollment.

We will hypothesize three alternative sets of objectives which can be achieved through a differential tuition policy, that the Regents of the University of California might have concerning the income group composition of the University's student body. The three sets of objectives have the common characteristic that the University has no preference for subsidizing students from one group more than students from any other group. They will differ in the University's willingness to make sacrifices in terms of the size of total enrollment for movements toward

---

* Associate professor, School of Public Affairs, University of Minnesota. The research underlying this paper was done at the University of California, Berkeley, and the paper was partly written at the Institute for Defense Analyses. The author gratefully acknowledges the support of both institutions. He wishes to thank George F. Break, Charles J. Hitch, John E. Keller, and Earl R. Rolph for encouragement and guidance during the study, and George A. Akerlof, Douglas C. Dacy, R. Daniel Gardner, Joseph A. Pechman, Henry M. Peskin, Rolf Piekarz, T. Paul Schultz, and an anonymous referee for invaluable suggestions on earlier drafts of this paper.

[1] Most of this subsidy is received from the state of California. This paper will not address the question of how large the subsidy should be.

[2] The existing empirical work on the private demand for higher education which is based on observed behavior of students has been done by Richard Ostheimer, Robert Campbell and Barry Siegel.

equal student representation from all groups. These objectives will be maximized subject to the constraints given by the size of the University's total subsidy and each group's enrollment demand function. The corresponding enrollments and fees for each group will be presented.

Section I describes the optimization procedure by which a university policy maker can use information about private demand for higher education in order to achieve his objectives with respect to total enrollment and the income distribution of students, given a fixed amount of subsidy. The use of this procedure is illustrated with the situation of the University funds in 1967. Section II summarizes briefly some of the results of a separate study (Hoenack) on private demand for attendance at the University. Section III shows the use of these results to determine the fees which would be charged and the enrollments which would result if the University were to use the optimization procedure outlined. I will present the results for the three alternative hypotheses about willingness to trade size of total enrollment for movements toward equal representation from income groups, defined by parental incomes of high school seniors.

I. Illustration and Application of the Optimization Model

This section presents a general description of a procedure for using information about private demand for higher education to help achieve objectives concerning the size and composition of total enrollment, given a fixed total subsidy for college students. Then it explains a specific application of this procedure to the situation of the University in 1967.

Illustration of the Optimization Model

Figure 1 illustrates the optimization model. Assume that there are only two categories of students, say students from families having incomes above and below the median level of family income in the population, and that their enrollments are measured A_1 and A_2, respectively. Assume that costs of educating students in the two groups are equal, and that demand for higher education is a positive function of income.

The area O*ab* represents the set of possible combinations of enrollments which can be achieved with a given subsidy. The curve, *ab*, is referred to as the enrollment possibility function (*EPF*). Since the higher income students have a higher attendance demand, the value O*b* exceeds O*a*; that is, the total enrollment would be larger if the entire subsidy were allocated to higher income students than to lower income students.[3] Both groups are equally represented at point *g*, where the 45° line intersects the *EPF*. At point *e*, where the slope of the *EPF* is minus one, total enrollment is at its highest possible level. In the portion of the *EPF* represented by *be*, there is no choice problem because reallocations of the subsidy increase both equality and the level of total enrollment. In the portion of the *EPF* represented by *eg*, any increase in equality reduces enrollment.

The curves labeled 1, 2, and 3 represent

---

[3] The *EPF* curve will in fact not touch the axes because demand is positive at a price which exceeds the university's cost of financing a college education. In this case the university will be making a net profit by taxing additional students in this category, thereby adding to the funds available for other students.

The convexity of the *EPF* curve can be demonstrated by calculating the second derivative of the function. The demand curve for A_i is $p_i = F_i\ (A_i)$ where p_i is tuition. The subsidy per student is the difference between c_i, the cost per student, and the tuition charged in category i. Thus, cost$-$subsidy$=F_i\ (A_i)$, and the *EPF* function is defined by the constraint $\sum (c_i - p_i)\ A_i = \text{constant}$. The slope of the *EPF* function is then

$$\frac{dA_1}{dA_2} = -\frac{[c_2 - MR_2]}{[c_1 - MR_1]}$$

where *MR* is marginal revenue. The convexity of the function is then easily shown.

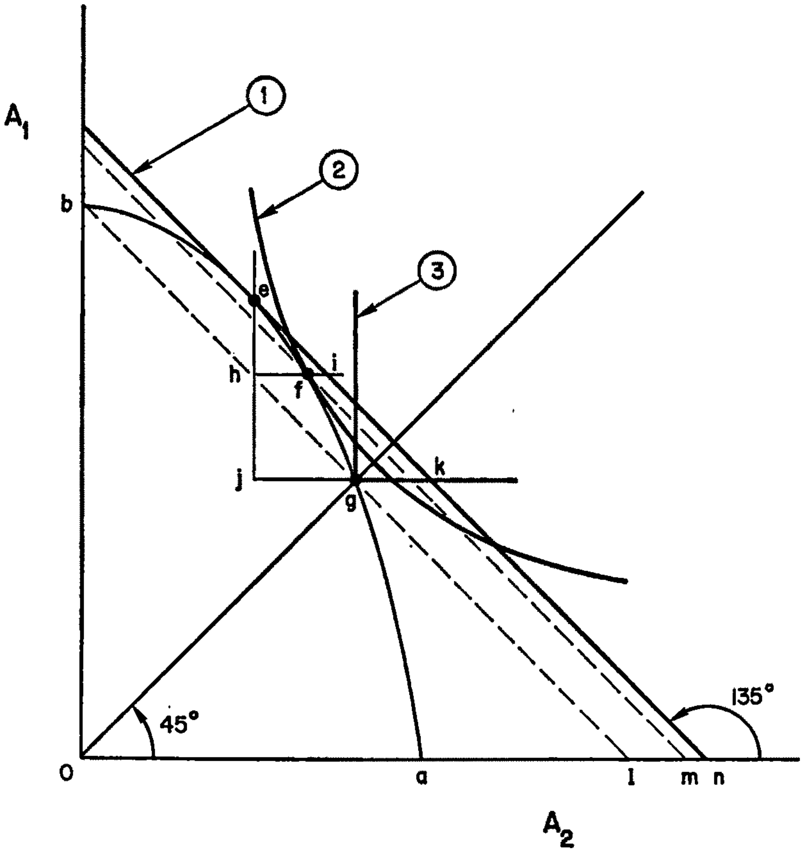


FIGURE 1

the indifference surfaces for three objective functions that might be used by policy makers to decide on the composition of the student body and total enrollment. With the first function the relative valuation of additional students in each income group does not vary with the relative level of enrollments from each group. While he has no preference for students in either group, the policy maker is concerned with enrollments. In the second function, the relative valuation of students from the two income groups depends on the relative enrollments from each group. When enrollments are unequal, the policy maker is willing to sacrifice more than one student from the group with higher enrollments for an additional student from the group with lower enrollments. This willingness to sacrifice size of total enrollment for movements towards equality increases with the degree of inequality. In the third function, only equal increases in enrollments from all groups are permissible. In this case, the policy maker's overriding concern is ensuring that the university's enrollments are drawn equally from each group, regardless of costs in terms of the size of total enrollment. Note that in the last two cases the policy maker is equally concerned if either group is disproportionately represented at the university.

The dashed lines in Figure 1 facilitate measurement of the sacrifice of total enrollment resulting from maximizing the second and third objective functions, instead of the first. The decrease in enrollments in either case equals the distance measured on either axis, of the dashed line from the solid line (the indifference surface for the first function).[4]

Application of the Model

For the purposes of this study all 1967 California high school graduates were divided into four equal sized groups determined by their parental incomes. These quartile income brackets represent the broadest available socioeconomic measure of the population served by the University. The reason for taking equal numbers in the population brackets is to facilitate the formulation and interpretation of possible objective functions concerning the size and composition of enrollments. While the objective functions show value for total enrollment and varying degrees of value for the composition of total enrollment, they show no preference for any particular category of students. Only the position of the constraint provided by the enrollment possibility function can lead the policy maker to enroll a relatively large number of students from any particular one of the equally sized groups. A_1 exceeds A_2 in the case where the policy maker maximizes the first and second objective functions. The reason is the relatively lower cost from his point of view of enrolling students from the more well-to-do median group because of this group's relatively high private demand for higher education.

A further consideration is that the University's admissions standards are stringent. The illustrative calculations presented here are based on the assumption that admissions standards will not be changed as part of a policy to change the composition of the student body. Approximately the upper 12-1/2 percent of California's graduating high school seniors are eligible students within each population group. In Section II, I show that student representation at the University from these subpopulations depends on the costs of attendance. In addition, a high school student is more likely to prepare himself to meet the eligibility requirements to attend the University the lower are the costs of attending. Therefore reallocation of the subsidy affects the relative sizes of the eligible subpopulations.[5]

It would be possible to extend this analysis and consider admitting some students from low income groups who would require remedial education to become University eligible. However, the particular set of calculations presented here is concerned only with University eligible students.

The following discussion is a formal presentation of the optimization model used.

The Constraint (enrollment possibility function)

The enrollment possibility function expresses the set of feasible levels of enrollments from the four quartile parental income groups, given the size of the subsidy. Let

TS = total subsidy given to students from all parental income quartiles
p_i = tuition paid by students from the ith parental income quartile
A_i = number of students from the ith parental income quartile
c_i = the average cost of the university per student from the ith parental income quartile

Then[6]

---

[4] To see that this is so, consider measurements along the horizontal axis. In the case of maximizing the second function, A_1 is reduced by *eh* and A_2 is increased by *hf*. Therefore, total enrollments are reduced by *fi* which of course equals *mn*. Similarly, enrollments are reduced by *ln* if the third function is maximized instead of the first.

[5] The University ordinarily attempts to redefine its eligibility requirements in order to maintain its eligibility pool in the top 12½ percent of California high school seniors.

[6] An extension of this relationship can assist the policy maker who is interested in the choices between quality of education and the numbers of students

(over)

$$TS = \sum_{i=1}^{4} (c_i - p_i)A_i \qquad i = 1, 2, 3, 4$$

The cost data in the relevant *EPF* for the University of California were based on the assumption that the c_i are the same for all parental income groups and that they do not vary within the relevant range.[7] The relationship between p_i and the A_i is the demand function of group i.

The Objective Functions

Let

α_i = the weight showing the relative importance of parental income quartile $i(\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha$ where α is constant)

A_i = the number of students from parental income quartile i

B = level of achievement of the policy maker's objectives

Objective Function 1

$$B = \sum_{i=1}^{4} \alpha A_i = \alpha \sum_{i=1}^{4} A_i$$

In this case the policy maker is concerned with the total level of enrollments. The composition of enrollments is not relevant to his objectives.

Objective Function 2

$$B = \prod_{i=1}^{4} A_i^{\alpha} = \left(\prod_{i=1}^{4} A_i\right)^{\alpha}$$

In this case the policy maker is concerned both with the level and the composition of enrollments. His willingness to sacrifice size of the enrollments for movements toward equal representation from the four quartile groups is proportional to the existing inequality of representation from the groups.

Objective Function 3

$$B = [\min (A_1, A_2, A_3, A_4)]$$

In this case the policy maker attaches no value to any increase in enrollments unless there is an equal increase in enrollments from all groups. Operationally this objective function was maximized by maximizing the first objective function subject to the constraints that enrollment from all quartile parental income brackets had to be equal.

Maximization of the Objective Functions

The following Lagrangean expression was used to maximize each objective function subject to the enrollment possibility function:

$$\begin{aligned} L = B(A_1, A_2, A_3, A_4) &+ \lambda[A_1(c - p_1) \\ &+ A_2(c - p_2) + A_3(c - p_3) \\ &+ A_4(c - p_4) - TS] \end{aligned} \tag{1}$$

The first-order conditions for the constrained maximization are:

$$\frac{\partial L}{\partial A_i} = \frac{\partial B}{\partial A_i} + \lambda \left[c - p_i - A_i \frac{\partial p_i}{\partial A_i} \right] = 0 \qquad (i = 1, 2, \ldots, 4) \tag{2}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{4} A_i(c - p_i) - TS = 0$$

Equations (2) are a system of five equations with nine unknowns: the four A_i's

---

educated. Assume that average cost per student (assumed to be the same for all types of students) is an effective index of quality. Also, assume that the college choices of students are influenced by quality. Then, $TS = \sum (c - p_i)A_i$ and $p_i = g^{(i)} (A_i, c)$ is the inverse of the demand functions. We then have $TS = \sum [c - g^{(i)} (A_i, c)]A_i$. One would expect that in some (low) range of values of c, increments of quality would bring about larger increments of p_i for given A_i. In this range the policy maker could simultaneously achieve larger enrollments and improved educational quality. Beyond some point, however, the p_i for given A_i would not increase enough to cover fully increases in c_i. In this range of c, the policy maker must choose between educational quality and size of total enrollments.

[7] Examples of possible differences in the c_i among groups of students are differences in instructional costs when students are grouped according to field of study and when students are grouped according to household characteristics, and remedial training programs for especially disadvantaged students.

and p_i's and λ. Each p_i was expressed in terms of its corresponding A_i in the appropriate demand function, and thus eliminated.[8] The system of five equations was then solved for the A_i's and λ. The objective maximizing values of each A_i are then achieved by setting tuition equal to its corresponding value of p_i in the ith parental income group's demand function. Following a brief discussion in Section II of how the demand functions were estimated, the results will be presented in Section III.

II. The Estimated Demand for College Attendance

A major objective in estimating demand for attendance at the University of California is to estimate how costs affect enrollments for each of the parental income groups. A time-series model would have been inappropriate for this task because the available data are inadequate.

A cross-section model was used in which the observations were based on several hundred California high schools. Since tuition does not vary across such a sample of high schools, it was necessary to base differences in costs on transportation expenses. The monetary cost of transportation was estimated on the basis of the cash and time costs[9] of driving a car to and from college. The high schools were divided into two categories: those within and not within commuting distance of a University campus.[10] In the case of individuals who live within commuting distance of a campus, the transportation costs are the daily costs of traveling from home to college.[11] For individuals living beyond commuting distance of a campus, the transportation costs are the costs of traveling home on vacations. In regard to the demand behavior of eligible individuals living within commuting distance of a campus, only a sample of about sixty high schools within commuting distance of the University campus at Los Angeles was used. The estimated demand function based on this sample was assumed to represent the behavior of eligible individuals living within commuting distance of all university campuses. Separate demand functions for eligible individuals not residing within commuting distance of a campus were estimated for each campus except for the San Diego and Santa Cruz campuses. The demand for attendance at these two campuses was assumed to be the same as for the other University campuses combined.

For the observation on each high school the following data were used: number of graduating seniors attending each campus of the University, number of graduating seniors eligible to attend these institutions, incomes of families residing in the high school's attendance zone (through matching census tract data with maps of high school attendance zones), and wage rate and unemployment data for the part of the state in which the high school is located.[12] In addition, use was made of a computer program provided by the California Division of Highways which calculated the minimum driving distance and time between each high school and each University campus.

---

[8] The variable p_i in this discussion is the positive or negative level of tuition. The direct cost variable in the demand function is the total expense of going to college including tuition, if any. When this cost variable is solved in terms of A_i, it is necessary to subtract a constant representing all costs other than tuition from the solution when it is entered in place of p_i in the system of equations. In 1967 a fee was charged at the University of California to cover certain non-instructional services. The *EPF* for the University as defined above was the same whether or not this incidental fee is taken into account and the discussion ignores this fee for simplicity of exposition.

[9] The assumed time costs underlying the estimates in Table 1 were based on data from labor market surveys provided by the California Department of Employment. The results are not very sensitive to alternative assumptions about the valuation of travel time.

[10] Maximum feasible commuting distance is defined using assumed valuation of travel time, travel expense, and frequency and length of trips, as the distance at which it would be equally expensive to live at home and commute to college as to live on campus.

[11] In the demand study (see Hoenack) it was assumed that any individual living within commuting distance of a University campus who chose to undergo the extra expense of attending a campus away from home would not be affected in his choice to attend a University campus if costs were to change within a range which encompasses the variation of costs considered in the calculations in Section III.

The following model was estimated. The jointly dependent variables were proportions of eligible high school seniors going on to attend each campus of the University of California. The independent variables were incomes of families in high school attendance zones, costs of attending each University campus and the nearest California State College and public junior college campuses, and wage and unemployment rates. (In the sample used for estimation of the attendance demand of eligible individuals living within commuting distance of U.C.L.A., there was no variation in costs of attending other University campuses or in wage and unemployment rates.) The remaining independent variables were interaction terms between each attendance cost and income. These interaction terms were used to estimate demand behavior separately for different income groups. For every estimated equation, the partial derivative of attendance proportion with respect to a change in the costs of attending all of the campuses was calculated for each income group. These partial derivatives were then weighted and summed to form the estimate of demand in each income group for attendance in the whole University system. Particular care was taken to make sure that the income variables used approximated the incomes of University eligible individuals. This was done by taking a sample of high schools for which both incomes in its surrounding census tracts and incomes of parents of University eligible students were known. These incomes were related to each other in a separate regression analysis. A detailed presentation of this demand study including the regression equations and the calculations made from them is found in my doctoral thesis.

The demand functions were estimated in two different ways in order to distinguish between short-run and long-run effects of changes in direct outlay costs. Lowering costs in the long run encourages students to become academically qualified and therefore raises the percentage of students eligible to attend the University of California, holding admissions standards constant. In the short run, however, a change in costs can only affect those students who have already qualified for entrance. The long-run demand for University attendance was obtained by adding to the short-run demand the estimated change in the number of eligible students resulting from a change in costs.

A logarithmic functional form was used in the estimation of the demand equations because it fit the data the best. Thus the estimated demand curves are steeper, the higher the cost of attending a University campus.

On the basis of the estimated relationships of the demand for the attendance at the University, the short-run elasticity of demand for attendance is −.85. The elasticity varies from −1.12 for the lowest quartile of parental incomes of all 1967 California graduating high school seniors to −.71 for the highest income quartile.

California state colleges are close substitutes for attending the University. Therefore, enrollment response would depend on whether there are simultaneous

---

[12] These costs were separately included in the regression equations only to help ensure that the equations were correctly specified. Foregone earnings should not be lumped with out-of-pocket costs because they represent in addition to opportunity costs of college attendance a potentially greater income while in college due to part-time and summer employment. Since college attendance is positively associated with income, the income effect of a rise in foregone earnings tends to offset its negative opportunity cost effect.

TABLE 1—PREDICTED EFFECTS OF INDICATED CHANGES IN ANNUAL COSTS OF ATTENDING THE UNIVERSITY OF CALIFORNIA ON 1967–68 UNIVERSITY ENROLLMENTS OF CALIFORNIA RESIDENT FRESHMEN

Income Bracket	Changes in Annual Costs	Percentage Change in Enrollment			
		No Changes in Costs of Attending Other Institutions		Costs of Attending California State Colleges Change by 2/3 the Changes in Costs of Attending the University	
		No Effect of Costs on Eligibility	Costs of Attending College Affect Eligibility	No Effect of Costs on Eligibility	Costs of Attending College Affect Eligibility
$ 0,000–7,599	−400	30.04	36.78	17.87	24.61
	−100	6.87	8.24	4.08	5.45
	0	0	0	0	0
	+100	− 6.04	− 7.04	− 3.65	− 4.65
	+400	−23.03	−26.05	−13.78	−16.80
$14,500–& Over	−400	21.03	26.42	12.71	18.10
	−100	4.66	5.69	2.81	3.83
	0	0	0	0	0
	+100	− 3.81	− 4.47	− 2.57	− 3.23
	+400	−14.38	−16.06	− 8.90	−10.58
All Incomes	−400	24.31	30.24	14.51	20.44
	−100	5.46	6.62	3.26	4.43
	0	0	0	0	0
	+100	− 4.59	− 5.38	− 2.98	− 3.78
	+400	−17.45	−19.66	−10.67	−12.89

changes in costs of attending these other institutions. The elasticities were recomputed on the assumption that any percentage increase in direct outlay cost of attending the University is simultaneously matched by a two-thirds increase of that change in direct costs of attending the state colleges.[13] The recomputed elasticity of demand for freshman attendance at the University is −.51. Similar reductions in elasticities apply to all income quartiles on this assumption: The recomputed elasticity is −.68 for the lowest quartile and −.48 for the highest.

Table 1 presents the effects of $100 and $400 changes in direct outlay costs on proportions of eligible California high school seniors who attend the University of California for the lowest, highest, and all family income quartiles. The table shows short-run and long-run estimates.[14]

III. The Results

The system of five equations corresponding to each of the three objective functions was solved numerically. The results are presented in Table 2, under the alternative assumptions that costs of attending institutions other than the University of California are unchanged, and that the costs for any income group of California State Colleges change by two-thirds of any alter-

---

[13] The assumption of a two-thirds simultaneous change is used for illustrative purposes here; the figure was chosen because at one time it had policy relevance in California.

[14] The University is legally required to adjust periodically its admissions standards in such a way that the proportion of high school students who are eligible is constant. The long-run results presented here are based on the assumption that the aggregate number of eligibles is not held constant. Other long-run estimates were made on the basis of the assumption that eligibility standards are redefined so that the total number of eligibles is constant.

TABLE 2—ENROLLMENTS OF 1967–68 RESIDENT FRESHMEN AND CORRESPONDING TUITION CHARGES FOR ALTERNATIVE OBJECTIVES

A. Assumption: No Changes in Costs of Attending Other Institutions

Objective Function	Enrollment in Each Quartile Income Bracket								
	Quartile 1 Under $7,599		Quartile 2 $7,600–$10,499		Quartile 3 $10,500–$14,499		Quartile 4 $14,500+Over		All Income Groups
	A_1	P_1	A_2	P_2	A_3	P_3	A_4	P_4	A
Status Quo	2136	0	2577	0	4083	0	6157	0	14,953
Objective Function 1	2474	− 155	2836	− 84	4193	+ 21[a]	5863	+ 168	15,366
Objective Function 2	3035	− 488	3246	−307	4000	+ 97	4866	+ 474	15,147
Objective Function 3	3587	− 817	3587	−493	3587	+259	3587	+ 865	14,348

B. Assumption: Positive and Negative Tuition Charges in State Colleges Equal Two-thirds the Charges in the University of California

Objective Function	Enrollment in Each Quartile Income Bracket								
	Quartile 1 Under $7,599		Quartile 2 $7,600–$10,499		Quartile 3 $10,500–$14,499		Quartile 4 $14,500+Over		All Income Groups
	A_1	P_1	A_2	P_2	A_3	P_3	A_4	P_4	A
Status Quo	2136	0	2577	0	4083	0	6157	0	14,953
Objective Function 1	2454	− 230	2814	−136	4176	+ 14[b]	5876	+ 215	15,320
Objective Function 2	2973	− 699	3195	−449	3997	+120	4940	+ 639	15,105
Objective Function 3	3533	−1205	3533	−726	3533	+393	3533	+1060	14,132

[a] Defined by quartile income ranges of spring 1967 high school graduates. Negative tuition charges represent grants-in-aid.

[b] Positive charge corresponding to enrollments which are close to status quo enrollment levels are due to partial linearization of the demand functions.

ation in the tuition at the University of California. Thus, the second assumption implies a broader program in which the state reallocates its entire subsidy for all college students rather than only that part of its subsidy given through the University of California. Table 3 shows the effects of the alternative policies on enrollment composition.

If the University of California desires to maximize its total enrollment (Objective Function 1), it will change tuition to increase enrollments from the lowest two income quartiles by about 10 percent.

If the state's values are represented by the second objective function, it will allocate the subsidy to increase enrollments in the lowest two quartiles by 33 percent over the present levels. This allocation of the subsidy actually increases total enrollment by about 1 percent. The increase in the first and second quartiles is gained almost entirely at the expense of enrollment in the fourth quartile.

In the case of the third objective function, the state will reallocate its subsidy to increase enrollments in the lowest two quartiles by 50 percent, while decreasing enrollments in the highest two quartiles by approximately 30 percent. With the third objective function total enrollments are 4 percent lower than currently. In general, movements in the direction of equality are increasingly costly in terms of total enrollment.

The two alternative assumptions in

TABLE 3—PERCENTAGES OF ALL SPRING 1967 CALIFORNIA HIGH SCHOOL GRADUATES WHO ATTEND THE UNIVERSITY OF CALIFORNIA UNDER ALTERNATIVE OBJECTIVE FUNCTIONS BY PARENTAL INCOME GROUP

Objective Function	A: Percentage Representation from the Population of Persons in Each Quartile Income Bracket				
	$0,000–7,599	$ 7,600–10,499	$10,500–14,499	$14,500–& Over	All Incomes
Status Quo	3.4	4.1	6.5	9.8	6.0
Objective Function 1	3.9	4.5	6.7	9.3	6.1
Objective Function 2	4.8	5.2	6.4	7.7	6.0
Objective Function 3	5.7	5.7	5.7	5.7	5.7
	B: Percentage Representation of Enrollment of Persons in Each Quartile Income Bracket				
	$0,000–7,599	$ 7,600–10,499	$10,500–14,499	$14,500–& Over	All Incomes
Status Quo	14.3	17.2	27.3	41.2	100
Objective Function 1	16.0	18.4	27.5	38.1	100
Objective Function 2	19.9	21.6	26.6	31.9	100
Objective Function 3	25.0	25.0	25.0	25.0	100

Table 2 do not differ substantially in their corresponding enrollments, but they do differ in the changes in tuition charges necessary to maximize the objective functions. The reason for the large differences in tuition charges is that enrollment demand for the University of California is less responsive to changes in costs when costs of attending other suitable institutions change in the same direction.

Tables 2 and 3 reveal an interesting characteristic of the present tuition policy of the University of California. Evidently, the constrained maximization of the first objective function leads to enrollment levels in all of the income groups which are closest to the corresponding status quo enrollments. Therefore, the present policy in California comes closest to maximizing the first of the three stated objective functions. It appears that unlike the illustrative functions used in this study which show no preference toward any income quartile, the existing implicit objective function of the University does show some preference for the upper two income quartiles. This situation does not result from any conscious bias favoring relatively well-to-do students but rather because of the nature of the private demand for attendance. In order to achieve an explicit set of objectives the University should base its fees on the private demand for college attendance.

REFERENCES

R. Campbell and B. N. Siegel, "Demand for Higher Education in the United States," *Amer. Econ. Rev.*, June 1967, *57*, 482–94.

S. A. Hoenack, "Private Demand for Higher Education in California," unpublished doctoral dissertation, Univ. Calif., Berkeley 1967.

R. H. Ostheimer, *Student Charges and Financing Higher Education*, New York 1953.

Discrimination by Waiting Time in Merit Goods

By D. NICHOLS, E. SMOLENSKY, AND T. N. TIDEMAN*

Perhaps the most ubiquitous of all urban problems is that the cities' public facilities—their roads, airports, shopping streets, license bureaus, schools, parks, beaches, pools, day care centers and public health clinics—are frequently crowded in ways that inflict time costs upon users. Waiting time does allocate public services, rationing them, as would money prices, according to the tastes, income and opportunity costs of consumers.[1] Time prices differ from money prices, however, since they appear relatively lower to persons with a lower money value of time. While such persons are likely to be considered more deserving, time prices have a defect: queues are a burden. It is alleged that some people, English housewives for example, enjoy a good wait. Despite such assertions, we will assume that queuing raises the cost of acquiring the good with which it is associated and that the burden from queuing is a deadweight loss. Time spent in a queue cannot be used productively.

The deadweight loss produced by a queue depends directly on the opportunity cost of time of those who wait. Thus when two individuals who value their time unequally wait in the same queue, they face different prices. This departure from the usual equilibrium conditions implies, in itself, that a queue of persons with different opportunity costs of time is inefficient. If trade were possible among persons who are waiting, or who might be paid to wait, this particular manifestation of inefficiency would disappear. Those with a low opportunity cost of time would resell to those with a high opportunity cost. Still, individual differences in the opportunity cost of time will affect the burden imposed by queues, because such differences determine who will wait and the length of the queue.

Money prices may be preferred to time prices because the revenues generated usually constitute an accurate signal. *Ceteris paribus*, it would be desirable to avoid the deadweight loss and to add to seller receipts. For these reasons, economists often recommend the imposition of user charges set equal to marginal social cost. However, a congestion charge in money is likely to be regressive in its effects, and several writers have agonized over whether the charge is justified simply because efficiency is increased in the process.[2] An alternative is to offer a public service at a wide range of money and time prices in a way that makes everyone better off. Our intention in this paper is to refocus the discussion of the efficacy of user

---

* Associate professor of economics, The University of Wisconsin, Madison; professor of economics, The University of Wisconsin, Madison; and assistant professor of economics, Harvard University. We wish to acknowledge the contributions of Samuel Morley, Jerome Rothenberg, Burton Weisbrod, and the participants in the 1968 Conference on Urban Public Expenditures where a preliminary version of this paper was presented. We are also grateful to that Conference and the National Institute of Mental Health for bearing a part of the costs of writing this paper. We were motivated to take up the problem of non-price rationing in the public sector after considering the provocative position taken by Julius Margolis (1968, p. 545).

[1] For a general discussion of time in the household production function see Gary Becker.

[2] See, for example, John Meyer et al. (pp. 334–41) and C. Sharp. Implicitly, these authors believe that there will not be any compensating redistribution.

charges on such an alternative program. At issue is not simply whether there ought to be a user charge in money at a single congested public facility, but rather how to achieve some constrained efficient allocation which is equitable.

Our single chain of argument will yield three major conclusions. First, we note that public services are frequently offered at a zero money price and then rationed by the waiting time required of recipients; furthermore, waiting time varies with the number of recipients. Since time is more equally distributed than money, this rationing device may be thought to be desirable because of equity considerations even though it is known to be economically inefficient. Since such equity considerations play no part in providing goods in the private sector, we conclude that public facilities are often congested for a reason in addition to those which lead to congestion in the private sector.

Our second principal conclusion is that queuing may be efficient. For public or private goods, queuing can be efficient if waiting by customers permits greater output. The efficient combination of queuing and money prices depends, of course, on the value of the customers' time. Queuing for public goods can also be efficient if there is a cost and a value to discriminating among recipients according to the opportunity costs of their time. The advantage of queues in this case stems from the fact that they enable us to charge different money prices to different groups without administrative cost. Facilities with higher money prices will have lower waiting times. A choice is thus offered the buyers which allows them to pay for the service with that combination of money and time which is cheapest for them. To low income buyers, combinations involving relatively higher time costs and lower money costs will be cheaper.

Finally, we conclude that the use of money prices to provide product differentiation may simultaneously improve equity and efficiency. We examine some equity issues that are inherent in alternative schemes for dealing with congestion and show that in some cases equity is improved in one dimension while it is worsened in others. We conclude by examining those characteristics of the social welfare function which must be known before one can unambiguously recommend the use of many money prices to partition the market for some otherwise homogeneous government service. While we restrict this analysis to differentiation by money price, we urge that the public sector consciously consider varying its conditions of sale along many dimensions. The chain of argument is also important because it leads to interesting questions, each worth considering in its own right quite separately from its relationship to the others. For example: Is the emphasis upon the need for marginal cost pricing in the public sector misplaced? Does the greater opportunity cost of time of the rich throw them into the private sector while leaving the poor in the public sector for selected consumption goods? Why do we provide merit goods and how do we determine the optimal capacity at a public facility providing that service? Does the reason for providing a merit good suggest the terms under which the good ought to be distributed? More generally, which rules for distributing merit goods are fair and which are not? We suggest a framework in which these questions arise naturally and we provide answers to some of them.

I. Congestion and Price Differentiation

Differentiating Products by Money Price

In the private sector, one way in which products are differentiated is by the time required to purchase them. One can spend time searching out merchandise at a discount store, examining it, and waiting in line at the cashier. Alternatively, the identical commodity can be bought rapidly at a retail shop with the assistance of a clerk. The good will be more expensive at the shop, of course, since it costs the shopkeeper money to save the buyer's time. If competition prevails in the retailing industry, profits will be zero for both discount stores and retail shops. Customers with a high opportunity cost of time will prefer the shops while those with lower costs will use the discount stores. The equilibrium number of shops relative to discount stores will depend on the technical ability of stores to substitute time for money and on the distribution of buyers according to the costs of time. Assume that the only difference among the firms in an industry is the amount of time customers must spend purchasing their output. Competitive equilibrium in such a differentiated industry will have two requirements: First the profit rate in each productive process must be zero in the long run; second, each buyer must patronize that supplier which sells the commodity most cheaply, where the purchase price consists of the money price plus the value of time spent making the purchase.

To be in equilibrium the buyer must solve the following problem. Many ways exist to buy a commodity, some of which have high money costs but low time costs while others have high time costs and low money costs. A continuous frontier of such possibilities, *FF*, has been drawn in Figure 1. The buyer must choose that point from *FF* which minimizes his total cost. Following Becker, we assume the cost of the buyer's time to be his wage rate, and draw *AB* such that *AO/BO* is the buyer's wage.[3] The minimum cost point is *E*. Buyers with higher wage rates would prefer to pay with more money and less time. As the frontier is drawn, the technology of retailing is such that the store can save time for the buyer at some cost to itself. Buyers with high wage rates find it worth their while to save that time and pay the higher money price.

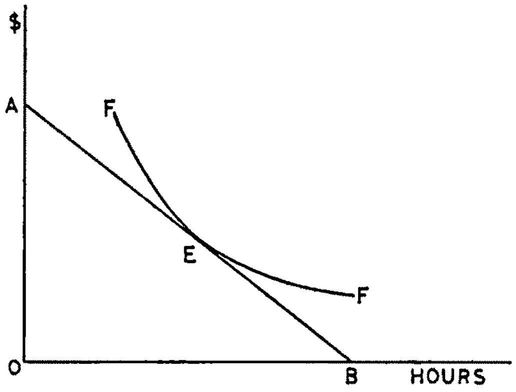


FIGURE 1

For producers, equilibrium requires that all points on *FF* yield zero profit.[4] Free entry guarantees this result in the long run. As long as average cost curves are U-shaped, an efficient competitive equilibrium results. This equilibrium is Pareto-efficient since the problem posed here is no different from the standard case of a firm in pure competition deciding what product to produce. Similarly, the individual's maximizing process is the usual one. We can view the purchase of each commodity as an activity with diminishing returns to labor. For an individual to maximize the return to his labor, his marginal minute in each activity must yield the same reward. If he faces a constant wage rate in one market, he must take part in all other activities until the marginal product of labor equals that wage rate in each activity. Thus the solution we have described is merely a special case of the general competitive solution. Its efficiency depends on conditions which are well known.[5]

---

[3] For *AEB* to be a straight line, we need to assume that the buyer faces a constant wage at which he may sell any amount of labor he chooses.

[4] If part of the frontier is non-convex, those points will not be observed as they represent inefficient processes.

To illustrate the gains to buyers that result from differentiating a product by money price where previously differentiation had not existed, consider an example in which the money price is initially zero (to represent the conventional public provision of a service). The commodity is offered subsequently at both a zero and a positive price. The initial situation is represented by *A* on Figure 2. At the zero money price, its use is rationed by the *OA* man-hours in waiting time it costs to acquire it. Later the product is also offered at a second facility at money price *OB*. If no congestion resulted at the additional facility and therefore it took no time to buy the new commodity, it would be bought by all those whose wage rates exceed *OB*/*OA*. A more general result would involve some congestion, with point *C* ultimately describing the cost of the commodity at the new facility. Since we are assuming capacity unchanged at the old facility, the demand withdrawn from it would result in a new time price such as *E*. Those whose wages exceed *DC*/*DE* would find it cheaper to purchase at *C*; others would purchase at *E*. Thus from the buyers' viewpoint, differentiation by price which involves increasing money prices can lower the total cost of acquiring a service for *all* consumers, provided that capacity has been added.

It is also possible for providers to prefer differentiation even with increased capacity. With the added capacity, total costs to the government increase. The new buyers brought into the government facilities by option *C* and those who switch from option *A* to option *C* because it is a cheaper option, generate revenue for the government. It is an empirical question whether the revenue offsets the increased costs. It is at least possible. In the case where net costs increase, some social decision criterion must be consulted to see if the extra benefit is worth the extra cost.

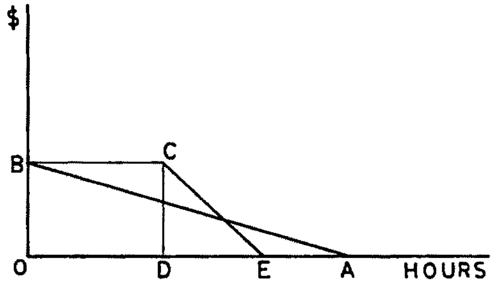


FIGURE 2

Each additional point which might be added to Figure 2 would entail a set of calculations like that above. In the limit a continuum of money prices would be created. Varying queues would exist with the longest queue associated with the lowest money price.

Fluctuations in demand are an important source of the money-time trade off. Consider a group of privately owned facilities using the same technology to produce a product which is differentiated only by the money price charged. Assume that demand fluctuates through the day and that it is administratively impractical to vary money prices as demand varies, so that queues form from time to time. Where the money price is low, congestion is more frequent and more severe, so that the average amount of time necessary to make a purchase is higher. Where the money price is high, there will be less congestion on the average and a lower time price. Thus fluctuations in demand can produce a frontier like *FF* in Figure 1, simply because buyers respond to different money prices. If zero profits are still to exist at all points on the frontier, there must be some fixed factor which leads to higher costs when the number of buyers is small. This would result, for example, from the existence of capacity that went unused at non-peak times. Buyers who wish to reduce the likelihood of queues must pay the costs of capacity which is not needed at non-peak times.[6]

---

[5] Note that low wage buyers are not able to resell to high wage buyers because of the transaction costs involved. The ability to substitute cheap labor for expensive labor has already been exploited in the frontier, and is in fact, the reason for its very existence.

Our concern is with publicly provided goods, but nothing said so far uniquely applies to them. Nor can public goods be introduced at this point by assuming insignificant long-run marginal costs, for if long-run marginal costs were insignificant there would be no congestion problem. To provide a rationale for public action it will be sufficient to assume that the commodity being provided is a merit good, i.e., that there is some public benefit from each unit sold. If such public benefits exist, and if those benefits do not depend on who consumes the commodity, then the efficient prices to charge individuals are given by a uniform downward shift of *FF* by the amount of the public benefit. For some publicly provided goods, however, merit value is related to characteristics of the consumer. Consider health services, for example. The poorer a person is, the more willing the public is to provide him with health services. It is this desire to differentiate among consumers according to income which undoubtedly provides the most satisfactory rationale for queues in the public sector, even though the time costs that result involve a dead-weight loss.

II. Queues that Deliberately Discriminate Among Merit Good Recipients

If the money cost of waiting time increases with the wage rate, any commodity that is rationed by a queue will be more expensive to those with high wage rates than to others. When confronted with alternative combinations of money and time prices, those with high wage rates choose the offering with a high money price and a low time price, while those with low wages choose the reverse. Thus if the public wishes to subsidize the money cost of a commodity to those with low wage rates only, they may offer it to all with a low (perhaps zero) money price, but offer such a small amount that a substantial queue results. To the low wage people, the money cost of the queue is minimal and they will receive a substantial benefit due to the lower money price. The high wage people will find the costs of the queue greater than the value of the money subsidy and they will not use the commodity even though its money price is low. Thus queues can be used to discriminate among users according to the opportunity costs of time. Even though the queue has an inefficient aspect in that the time of those who pass through it might have been used to raise total output without adversely affecting the buyer, nevertheless it is efficient overall if the alternative costs of discriminating—an equally effective means test—are higher. A queue is a decentralized way to discriminate according to the opportunity costs of time; it allows low wage people to select themselves as recipients of the money subsidy. Of course, if any alternative means of discrimination is cheaper, the queue remains inefficient.

Since queues may be the most efficient means of discrimination for some purposes, it is useful to discuss the nature of the problem faced by the government when determining the optimal length of a queue of a non-tradeable, non-storable commodity.[7] Queue lengths are determined indirectly, of course, since the control variables are the money prices charged and the quantities provided per time unit. The responses of individuals to these prices and to the resulting queues determine their lengths. For a given set of money prices, queues can be reduced by increasing the quantities of the services available.

---

[6] This happens, for example, at supermarkets where product differentiation is effected by varying the number of cashiers employed. The more cashiers, the less often queues occur but the higher money prices must be.

[7] We have in mind here goods like visits to a health clinic, where the opportunity to reduce waiting time by buying more of the good each trip is virtually impossible.

The optimal quantity of the product to offer at any price is that amount at which the social cost of an additional unit just equals the social benefit. Comprehension of the relevant benefits and costs yields an understanding of the optimal solution.

Consider the problem of a government which wishes to subsidize the consumption of a commodity by low wage individuals and can offer a fixed subsidy to all potential consumers. For one consumer, the problem is simple. It is well known that such a subsidy must equal the value of the utility gained by other persons from the last unit of the commodity consumed by the individual. In Figure 3, *DD* represents an individual's demand curve for a commodity and it is known by the government. At each quantity, however, there are marginal external benefits to the general public which when added to the individual demand schedule produce the total marginal benefit curve *D′D′*. Given marginal production cost of *MC*, *AB* becomes the appropriate subsidy. With subsidy *AB*, the individual chooses to consume *OC* while he would choose *OE* in the absence of that subsidy.

If, as may be assumed, the same subsidy must be given to different individuals, there will be some welfare loss since the amount of public benefit, at the quantity he consumes, varies from one individual to another. For the individual pictured in Figure 4, the subsidy *AB* is not large enough to induce consumption of *OH*, the optimal amount, and a welfare loss represented by triangle *AGF* results. For some consumers, the subsidies will be too large

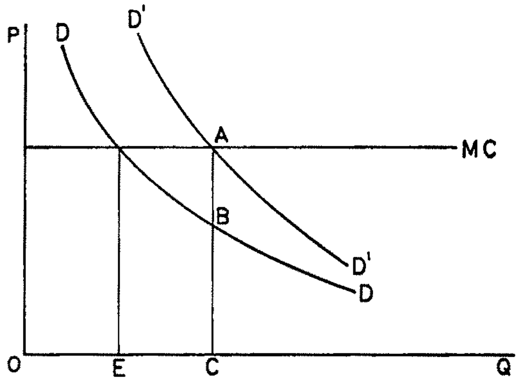


FIGURE 3

and corresponding triangles *AG′F′* will appear below the cost line, representing the fact that the value of the commodity to the individual plus its value to the government is less than its cost. If we continue to ignore the possibility of queues, the problem of the government is to select a subsidy scheme which minimizes the sum of triangles such as *AGF* and *AG′F′* over all individuals.

The use of queues to discriminate among users adds an additional source of welfare loss to that already represented by the triangles. This follows from the assumption that the opportunity cost of time is the wage rate, for by that assumption buyers would be indifferent between spending their time in the queue and paying out

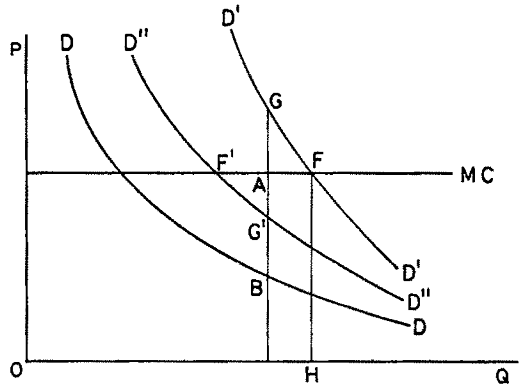


FIGURE 4

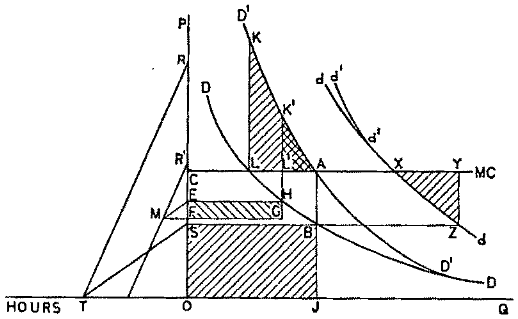


FIGURE 5

of the wages earned at their productive jobs. Nevertheless, the effect of the queues on the subsidies received by different individuals may reduce the total welfare loss. Queuing may lower the subsidy to those with high wages and raise it to low wage people. If this is the desired effect, it is possible that a welfare gain can result despite the additional deadweight loss implicit in the queue.

In Figure 5, the private demands of a low wage and a high wage individual are represented by DD and dd, respectively, while $D'D'$ and $d'd'$ represent private plus public demands for the same individuals. Note that the appropriate subsidies are AB and zero. If no subsidy is offered, the welfare loss is triangle AKL. If both are offered subsidy AB, the loss is triangle XYZ. Suppose that there is no uniform subsidy, and instead, the product is provided at a zero money price in the public sector and MC in the private sector, and a queue results in the public sector which requires an individual to spend OT man-hours to buy each unit. The individual's perception of the subsidized price will then be the money value of the waiting time. If the wage of the low wage person is OS/OT and that of the high wage person is OR/OT, then the money cost of the commodity in the public sector to the low wage person will be OS, while it will be OR to the high wage person. The time costs for the high wage person will exceed MC, the cost of the commodity in the private sector, so that the low wage person would buy from the public sector, while the high wage person would buy from the private sector. The only welfare loss results from the queue, a loss of $OJBS$, the value paid by consumers in waiting time that might, without additional cost to the buyer, have been used in production.

As drawn, $OJBS$ is very large; an alternative pricing device exists which reduces this loss substantially. If the public sector charges a money price OF and offers a quantity that results in a queue which requires FM man-hours to acquire the commodity, the high wage person will still use the private sector although the price to him of the service in the public sector has fallen substantially, i.e., from OR to OR', as long as the total cost to him, OR', exceeds the marginal cost of the alternative, OC. The low wage person now pays OF' in money plus FM in time for each unit and consumes EH units. The total waste is $EFGH$ plus $AK'L'$. A queue that minimizes the sum of this rectangle and triangle would be the most efficient scheme the government could offer when the costs of separating low from high wage people in any other way exceed that minimum sum. When the opportunity costs of individuals lie along a continuum, the government must select quantities of service to offer at various money prices in such a way as to minimize the sum of the rectangles and triangles over all individuals. That is, the sum of the wastes from standing in queues and from having different citizens consume too much or too little of a specific commodity must be minimized. In formal mathematical terms, this is a very messy problem; no significant solutions appear readily at hand.

A solution to a somewhat similar problem has been presented by William Vickrey. He discusses the provision of a nonmerit good (highway services) with time-cost related to the number of users and shows that if there are constant returns to scale, competition will produce a continuum of money price-time price pairings in which the price that each person pays will cover the congestion costs imposed on others. Our problem is different and more difficult in that congestion serves as a suboptimizing substitute for a means test rather than a way of making services available to more persons.

Differentiation of price may serve not only to vary the time cost to users, but also to vary other qualities of services. An obvious example would be the consumption of space at beaches, pools, and parks. Differentiating by price would leave some facilities less crowded than others, and allow thereby for differences in taste. One consequence might be to simply differentiate consumers by income class, which in at least some instances would be undesirable. Not all forms of product differentiation by money price are desirable, and we turn to a general consideration of their equity consequences in the next section. We conclude here by noting that examples of product differentiation can now be found in the public sector. Burton Weisbrod tells of a first-rate example of what is generally required. In San Juan there are "express" busses which run along the same routes as "local" busses and both are required to stop at the same places if their patrons demand it. Expresses, however, carry a higher price which tolls off customers and makes the express bus the more rapid travel mode. In this example the waiting time does not represent a deadweight loss since it is one of the necessary inputs for transportation, and is an example of Vickrey's model.

III. Equity and Money Price

The hard equity questions have been side-stepped until now since we have implicitly posited the existence of some social welfare function by specifying the public value attributed to each additional unit of consumption by an individual. Regardless of the form of that welfare function, certain systematic redistributions are implicit in any scheme involving the introduction of money prices into the public sector when they had not existed previously.

There will be a high but not perfect correlation between income and the opportunity costs of time. If we wish to treat those with equal income equally, we will find that the use of queues encourages too much consumption by those with low wages who have sources of non-wage income. And, of course, income may not separate those whom we wish to subsidize from those whom we do not wish to subsidize. Schemes which differentiate beneficiaries according to the opportunity costs of time are inappropriate if the society wishes to differentiate by other standards. The number and age of children, condition of health, or level of education may all affect the degree to which there is a public interest in enhancing the consumption of an individual, either in general or of a specific service. To the extent that these factors are present, queues will be an inefficient device for giving effect to such public interests.

Equity among income classes is also affected by the range of available prices. Suppose that a visit to a doctor is available at $0 plus 2-1/2 hours at the public clinic or $5 plus 1/2 hour in a doctor's office. The effective prices in money (the sum of money prices and time prices converted to money) for persons with different opportunity costs of time are shown by the solid line in Figure 6. Persons with time worth less than $2.50/hour use the public sector and pay 2-1/2 hours, while those with time worth more than $2.50/hour use the private sector and pay $5.00 plus 1/2 hour. Now suppose that an additional price is

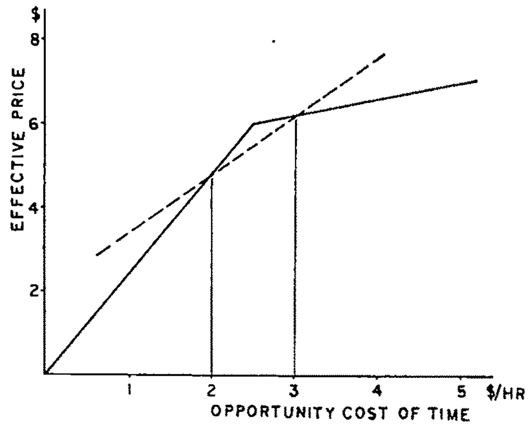


FIGURE 6

offered at the public clinic: $2.00 plus 1-1/2 hours. (This could be accomplished at a single facility by allowing each patient to specify the line he wished to join, and then calling patients in such an order as to preserve the relative lengths of the lines. The doctors would not need to know which line a patient came from.) The new effective prices, expressed in money, are shown by the dashed line in Figure 6. The beneficiaries are persons with opportunity costs of time in the range of $2/hour to $3/hour. The greatest benefits accrue to persons with time worth $2.50/hour.

The preceding discussion requires that the zero-priced facility continue to have the same time price. This is accomplished by adding an appropriate amount of capacity. If we were to merely institute money charges at some facilities where none existed previously, the queue lengths at the zero-priced facilities, serving the very poorest people, will increase. By varying capacity, any desired time price can be achieved.

It is difficult to set constraints on what constitutes equity. Nevertheless, it would be surprising if the corners of Figure 6 were consistent with maximization of any social criterion. Equity would seem to require a continuous variation in the nearly equal treatment of near equals, but a smooth schedule in Figure 6 would require a continuum of money-time price pairs, which might be prohibitively expensive. The practical optimum almost certainly involves undertaking some expense for the sake of greater continuity.

If we relax the assumption that the opportunity cost of time is measured by the marginal product of labor in the market, we can say little about equity. Relaxing the assumption does raise a pertinent question about equity, however, when the chosen policy is to expand the set of money price-waiting time pairings. Suppose, for example, that two individuals have the same opportunity cost of time in the labor market. For one of these individuals, the opportunity cost of time is indeed his marginal product of labor as valued in the market place. For the other, the opportunity cost of time is in performing motherly duties which she values at more than the market value of her time, so that she earns no wages. Introducing the set of money price-waiting time pairings benefits the worker but not the mother, who values her time so highly that she continues to pay the high money price.

Current practice which, by and large, offers public services on a first come, first served basis at a single money price (usually zero) poses its own problem in equity. As an appropriate test of an equitable rule we offer the following: *ex post* is the distribution of recipients of a service a random draw from the client population along each relevant dimension. That is, are there any systematic variations in characteristics between recipients and non-recipients which are arbitrary with respect to the purposes of the program? (See Morris Ginsberg, ch. 2.) Our expectation, which must still be empirically verified, is that current practice would not pass this test. Current practice imposes an arbitrary distinction: consumers will be differentiated from non-consumers by the opportunity costs of their time. The missing money price-waiting time pairings may even serve to discourage work effort from those at the margin between unemployment and employment at low wages. The value of waiting time may exceed the marginal product of labor at a full-time job. Casual employment may yield a higher total utility than the somewhat higher money income from full time employment for the poor hypochondriac—and hypochondria is a poverty-linked characteristic.

In summary, our policy proposal is to increase choice in the public sector through variation in money prices, which may be considered a form of third-degree price discrimination.[8] This proposal is not without equity problems. But without an explicit social welfare function on the one hand and a precise congestion function and production function for the service on the other, these conflicts are not resolvable.

Introducing Money Prices into the Public Sector: Some More General Issues

Failure to provide many alternative price-time pairings is not a widespread problem. Not only does it arise on a quite limited number of publicly provided goods, but it also probably affects the welfare of only a small portion of the income distribution. Many public facilities carry only one money price, usually zero, but the private sector supplies closely substitutable services at alternative price-time pairings. The two extremes of the income distribution are, therefore, probably getting the appropriate choices. The pairings offered, however, have a sharp discontinuity between the zero money price at which the good is fully subsidized by general taxes, and the minimum feasible money-time price pairing which just yields normal profits. Partial subsidies, with some money user charges, are lacking.

Partial subsidies have applications beyond those instances in which congestion is manifested by queues. Choice seems to be unduly restricted over the whole range of public service. One can go to the health clinic at zero money prices or the private doctor, the municipal golf course at zero money prices or the country club, the library at zero money prices or the second-hand book seller, the purely public or the purely private elementary school. The larger the public subsidy to any congested public service, the sharper will be the discontinuity in the price pairings offered. The general effect is to serve poorly those with a low but positive marginal product of labor.

Alternative price pairings for the same basic goods do exist throughout each metropolitan area. One source of these variations is residential segregation by income class. The charity health clinic that caters to the needs of domestics in high income areas is not likely to be congested. For other reasons, the public schools in the high income areas are not likely to be congested either. Political boundaries within the metropolitan area serve to permit individuals to collect according to income and their taste for public services, thereby producing a mix of money and time prices, but the commodity also varies.[9]

---

[8] Strictly speaking, third-degree price discrimination is an attribute of monopoly and an exercise of monopoly power. "A third degree would obtain if the monopolist were able to distinguish among his customers *n* different groups, separated from one another more or less by some practicable mark, and could charge a separate monopoly price to the members of each group" (A. C. Pigou, p. 279). We do not expect the government to exercise any monopoly power it may have simply to increase money receipts for its own sake.

[9] Charles Tiebout and Margolis (1957). However, as Paul Samuelson has pointed out, variance in income within neighborhoods reduces the effectiveness of product differentiation among municipalities in a metropolitan area (p. 377).

IV. Conclusions

The issues posed by this paper have not been completely resolved. We have done no more than highlight some relatively neglected facets of the problem of congestion at public facilities.

Because of the peak load problem, it is often efficient to have queues at both private and public facilities if the cost of varying prices exceeds the dead-weight loss of the queue. In addition, public services that are provided below cost often appear to be rationed through the use of queues. Such a rationing device is efficient only if the alternative forms of rationing are more costly than the dead-weight loss implicit in the existence of the queue. Queues are effective rationing devices because they impose a charge on the users in waiting time. Since the opportunity costs of time vary across people, the money cost of the queue will vary as well and for many people the costs of waiting will exceed the price at which the service can be bought in the private sector. Those with a high opportunity cost of time will find the money price in the private sector to be lower than the time price charged in the public sector. Thus the use of a queue rations a service exactly as if a money price were charged that varied directly with one's wage rate. Since the public sector often wishes to subsidize commodities in a manner that varies negatively with wage rates, queues can be an efficient device for singling out those it is desired to assist. This is true when alternative costs of discriminating exceed the dead-weight loss implicit in the queue.

In some cases, it may be desirable to charge many different money prices for the identical publicly subsidized commodity. Queues of different lengths will form with the shortest queues occurring at the facilities with the highest money prices. Individuals will then have a choice of paying for a commodity with various combinations of money and time, each choosing that combination which is cheapest for him. There may be substantial efficiency gains to be had from such differentiation.

Our proposal may produce serious equity problems that cannot be overcome.[10] Even if only the poor benefit, those with higher money income may benefit relatively more than those with lower incomes. If equity means the same treatment for all persons, it may not be possible to improve social welfare by increasing the number of money-time pairings. If, however, unequal treatment of unequals is equitable, which seems much more reasonable, then there are unexploited possibilities for improving social welfare. If offering the relatively better off among the poor services at both a smaller subsidy per capita than other poor and a smaller congestion cost is equitable, for example, then there should be a substantial increase in the set of the money-time price pairings offered the poor.

Taken together, the public and private sectors provide substitutable commodities at many alternative money-time price pairings. Those with a high opportunity cost of time will choose from the private sector, and the poor will choose from the public sector. Segmentation of the market will not extend to its technically feasible limits, however, unless governments offer income-in-kind at varying money prices. We think we have shown that there may be a high payoff in increased social welfare to ingeniously conceived expansions in the number of waiting time-money price pairings in the public sector. By extension, the payoff to increased welfare may also be extended by differentiating product along dimensions other than the money and time.

---

[10] The problems of extending choices on the supply side, thereby foregoing economies of scale, have not been discussed here.

REFERENCES

G. S. Becker, "A Theory of the Allocation of Time," *Econ. J.*, Sept. 1965, *75*, 493–517.

M. Ginsberg, *On Justice in Society*, Baltimore 1965.

M. B. Johnson, "On the Economics of Road Congestion," *Econometrica*, Jan.-Apr. 1964, *32*, 137–50.

J. Margolis, "Municipal Fiscal Structure in a Metropolitan Region," *J. Polit. Econ.*, June 1957, *65*, 225–36.

———, "The Demand for Urban Public Services," in H. S. Perloff and L. Wingo, Jr., eds., *Issues in Urban Economics*, Baltimore 1968, 527–65.

J. R. Meyer, J. F. Kain, and M. Wohl, *The Urban Transportation Problem*, Cambridge 1965.

A. C. Pigou, *The Economics of Welfare*, 4th ed., London 1960.

P. A. Samuelson, "Aspects of Public Expenditure Theories," reprinted in J. Stiglitz, ed., *The Collected Scientific Papers of Paul A. Samuelson*, Cambridge 1966.

C. Sharp, "Congestion and Welfare—An Examination of the Case for a Congestion Tax," *Econ. J.*, Dec. 1966, *76*, 806–17.

C. M. Tiebout, "The Pure Theory of Local Public Expenditure," *J. Polit. Econ.*, Oct. 1956, *64*, 416–24.

W. Vickrey, "Congestion Charges and Welfare," *J. Transp. Econ. Policy*, Jan. 1968, *2*, 107–18.

Optimal Mechanisms for Income Transfer

By RICHARD J. ZECKHAUSER*

The poor receive a sub-optimal amount of material goods through the economic processes of our society. Poor and non-poor alike seek to institute mechanisms that give something to the poor beyond what they get from the natural workings of the "competitive" system. That the non-poor share in this desire indicates that an externality is generated when the economic welfare of the poor is improved.

On a collective basis, the non-poor are willing to pay for the provision of this externality through transfer programs that benefit the poor. In the traditional terminology of externalities, a transfer program leads to an increase in the level of production of a good, in this case the economic welfare of the poor, which generates an externality.

This externality goes to all the non-poor. Thus, taking the non-poor to be the relevant group, this externality functions as a public good.[1] Once it is provided for one non-poor man, it is made available to benefit each non-poor man whether or not he contributed to its provision. Exclusion is not possible. As is usual with public goods, if individuals act in isolation a sub-optimal level of the good will be provided. That is why redistributive schemes of any magnitude are usually carried out by some collective body.[2]

In this paper, I am interested in the transfer program that would be chosen by a representative citizen who took into account that he would have to pay his proportional share of any collective effort. I assume that the citizen acts as a utility maximizer. His utility function is taken as given, so I need not be concerned with the moral or ethical values that underlie its structure. It is not of consequence here, whether the externality that enters his utility function is motivated by guilt, a sense of justice, or the desire to maintain a stable society.[3] What is important is the way a non-poor man benefits from improvements in the material well-being of the poor, the way the externality enters his utility function.

The elusive social welfare function is not dealt with here, nor are any other criteria relating to the welfare of the total society. All relevant welfare judgments are made on an individual basis by the participants in the society. It is probable, as I have suggested, that the providers of the scheme would choose to act collectively.

I conduct the analysis with the aid of a simple two-party model. One party is the representative poor man. The other is the representative citizen; he is assumed to be non-poor.[4] The structure of the model is given by the assumed form of the representative individuals' utility functions and the permissible forms of income transfer schemes. After establishing this structure, I first examine the efficiency of negative income tax schemes. Then I turn to the major investigation, the problem of deriving the optimal income transfer scheme.

---

* Associate professor of political economy, Harvard University. I am indebted to a referee and the managing editor for many suggestions which have improved the paper. A preliminary and more detailed version of this paper was circulated in 1968.

[1] See James Buchanan for a recent discussion of the strategic elements of interactive situations in which groups provide themselves with public goods.

[2] Private charity cannot be relied upon to lead to a significant level of transfers to the poor unless there are substantial private benefits that return to the individual giver. Some individuals derive great inner satisfaction from giving, but for many others something more is needed. To this end, churches offer salvation to donors, universities name buildings for them, and the United Fund gives them stickers for their windows. The federal government encourages private charity through tax deductions.

[3] One need not be a Marxist to allege that poverty and societal stability do not go hand in hand. "Poverty is the parent of revolution and crime." Aristotle, *Politics*, Book II.

Quite obviously the optimal scheme in any particular circumstance will depend upon the parameters of the model. However, it is evident that certain predominant characteristics of optimal schemes can be discovered. To facilitate exposition, and indeed to make the optimization process computationally feasible, I sometimes find it helpful to employ numerical examples; in each case, the reader should understand that the specific functions and values of the parameters are selected to be illustrative, and that the major characteristics of the results can be generalized.

I. The Formulation of the Problem

The representative citizen asks himself the central question of this paper. *How should assistance programs to the poor be structured so as to maximize the utility function of the representative citizen?* I assume that considerations relating to assistance programs can be examined apart from the other decisions and allocations of the society, that the maximization of a subsidiary utility function relating to this single area will be consistent with the maximization of the representative citizen's comprehensive utility function.

My representative citizen, hereafter called A, has a utility function that includes three variables: The first variable, y, is the poor man's annual income after all transfers have been made. This variable, like all other money magnitudes in this paper, is measured in dollar units.[5] Assume for the present that A does not care about the poor man's expenditure pattern, that the poor man can best determine his own optimal consumption allocations, and that, other things being equal, the pattern that is best for the poor man is best for A. The second variable, h, is the number of hours the poor man works in a year. The third variable, c, is the annual dollar cost of the assistance program per poor man. The number of poor men in the society is known. Indirectly, then, c also gives the total annual cost of the program for the whole society. The total transfer to the poor, or perhaps A's share of it, may enter more directly into A's utility function. Given that these cost magnitudes are linear relations of one another, any one of them may be used as an argument of A's utility function. I use cost per poor man.

My representative poor man, hereafter called P, has a utility function that includes only the first two variables: his income and the number of hours he works. The utility functions for A and P are, respectively

$$(1) \quad {}_AU = A(y, h, c) \quad \text{and} \quad {}_PU = P(y, h)$$

Both functions are assumed to have positive partial derivatives with respect to y. The partial of A with respect to c is negative. It is not possible to prespecify the signs of the partials with respect to h for all values of h. Over the relevant range, however, we would expect the partial of P with respect to h to be negative. This means that P will choose to work fewer hours if he loses no income as a result. The partial of A with respect to h will be positive, at least until h takes on a very substantial value. Up through this point, other things being equal, A would like to see a reduction in P's leisure time.[6]

---

[4] In some societies the poor will exercise influence over the structure of assistance programs. In such societies the representative citizen is an amalgam of a poor and a non-poor man.

[5] To avoid the complications that accompany attempts to deal with incommensurate units, units of measurement are frequently omitted in this paper. Thus y here is given the value of 500, not \$500.

A's preference in this regard might reflect a paternalistic attitude. For P's sake, A might wish to provide him with the motivation, self-esteem, and beneficial experience that come from working at a job. Then again, A's preference may merely indicate his selfish desires. He may believe that additional work for the poor will reduce social unrest.

The major question of this paper, remember, is how should the representative citizen design assistance programs to the poor so as to maximize his own utility function. There is no general answer to this question. However, if we restrict the scope of our inquiry and specify the structure of the interactive situation, we can gain some insights. One possibility would be to treat this as a problem of trade in valued goods. A gives money to P in return for which P agrees to work some specified amount. The amount of the gift plus P's earnings would be the total of P's income. With full cooperation A and P can easily arrive at a Pareto optimal point. However, the type of cooperation needed to reach such an outcome might be impossible to achieve. Indeed, contractual arrangements of this nature might be considered unattractive.

Assistance programs are usually formulated on an impersonal basis. The transferring agent or agency, the representative citizen in this model, first draws up a general assistance program. The individual poor man then responds to the program in the way that is best for him. In terms of the model, A and P are engaged in a nonsimultaneous move, nonzero sum game. In the absence of cooperation, the situation has aspects of interdependence familiar to us from the theory of leadership oligopoly. There one firm, the first mover, markets a quantity assuming that the second firm will maximize its profits with respect to the market situation. This is the type of strategic planning that the representative citizen follows when he draws up the mechanism for income transfer that is best for him.

The model works in the following fashion. *A first sets up an assistance program; P then selects the number of hours he wishes to work in order to maximize his utility function subject to that program; A selects the program that gives him the highest utility payoff when P maximizes with respect to that program.* This formulation will be the basis for the analysis that follows.

II. The Negative Income Tax Plan

Economists have proposed the negative income tax (*NIT*) as a promising approach to the problem of redistribution. (See Christopher Green, Robert Theobald, James Tobin (1966, 1967), and James Vadakin.) Under the plan a poor man receives a stipend from the government, and this stipend is reduced as his income increases but by less than the full amount of the increase. The reduction in stipend represents a tax by the government on earned income. By contrast, income maintenance plans cut the stipend by one dollar for every dollar of earnings up to the base level. They form the limiting case of the *NIT* plan with the marginal tax rate equal to 100 percent.

It is unfortunate that these transfer plans have acquired the label negative income tax. They are negative only in the aggregate sense and apply a positive marginal tax rate to earnings. In this paper I follow convention to the extent that I label tax plans that are negative in the aggregate as *NIT*'s. However, as I will discuss at length, it is most important to distinguish between plans that have negative marginal tax rates and ones, like those that have been widely discussed, that apply positive marginal tax rates to earnings. Hereafter, in this paper, unless otherwise specified, the term "tax rate" refers to the marginal tax rate on earnings.

---

[6] Beyond some point, of course, A would like to see P have more, not less, leisure time. It is with this consideration in mind that we establish such measures as overtime laws and child labor laws.

The advantage of the *NIT* over income maintenance plans is that it retains some of the incentive to work. (See Peter Diamond.) At any income level some amount of marginal earnings will be retained. Figure 1 compares a positive marginal tax rate *NIT* plan with an income maintenance plan with the same base stipend. The recipient's indifference map is superimposed on the figure so it is possible to determine his maximizing response (the number of hours he chooses to work) for each of the two transfer plans.

In Figure 1:

wage rate = slope of *OA*,
base stipend = *OB*,
marginal tax rate under *NIT* = slope of *OA* minus the slope of *BD*,[7]
cost of the plan = the vertical distance from a relevant point on the two-segment line *BDA* to line *OA*.

Notice that northwesterly movements across the figure take the recipient to higher indifference curves.

As illustrated, the recipient would choose not to work under the income maintenance plan. Point *B* is on a higher indifference curve than any point on *CA*. With the negative income tax, he will choose point *E* where the two-segment line *BDA* reaches the highest indifference curve.

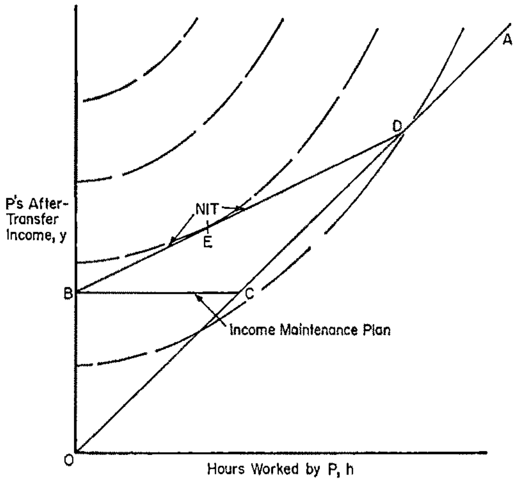


FIGURE 1. A NEGATIVE INCOME TAX PLAN COMPARED WITH AN INCOME MAINTENANCE PLAN

For the poor man, the *NIT* weakly dominates the income maintenance plan. For any amount of work it yields equal or greater income. Which plan is more costly to the government depends upon the amount that is worked under the two plans. The cost is given by the vertical distance from the income-hours point selected by the poor man to the wage-income line, *OA*. The more effective the *NIT* work incentive, the lower is its relative cost.[8]

---

[7] Beyond point *D* regular tax rates may apply to reduce income after stipend and taxes.

[8] The poor man might work more under the income maintenance plan than he would with the *NIT*, even though the partial derivative of his utility function with respect to work is everywhere negative. This will be the case if the best point for the poor man on *AC*:1) is on a higher indifference curve than point *B* (the assumed best point on *BC*) and therefore will be selected with the income maintenance plan; 2) is on a lower indifference curve than the best point on *BD* and therefore will not be selected with the *NIT*; and 3) lies to the right of the best point on *BD* and thus represents a greater work effort. Unless leisure is an inferior good, 3) will hold whenever 1) and 2) are satisfied.

The case in which these three conditions are satisfied is quite different from the case in which leisure is an inferior good, so that raising its price (decreasing the marginal tax rate on income) decreases the amount of leisure consumed (increases the work effort). If it were an inferior good, then as the price line pivoted upwards from *BC* to *BD*, higher utility levels would be accompanied by greater work effort. That is, the *NIT* would produce a greater work effort than the income main-

An Example with the Negative Income Tax

The possibilities A considers are all negative income tax plans with constant marginal tax rates. A selects the base stipend, s, and the tax rate, t. The transfer plan thus established determines P's after-transfer income, y, as a function of the hours he works, h, and his wage rate, w. This income equals the base stipend, plus P's wage income multiplied by one minus the tax rate. That is,

$$y = s + hw(1 - t) \tag{2}$$

Given the transfer plan, P picks h, the number of hours he wishes to work, in order to maximize his utility function, $P(y, h)$. In the terminology of oligopoly theory, P reacts to the s and t selected by A.[9] This response by P determines the cost of the plan to A. That cost, c, is the difference between P's after-transfer income and his wage income if he works h hours. In algebraic terms we have

$$c = [s + hw(1 - t)] - hw = s - hwt \tag{3}$$

A manipulates s and t to maximize his utility function, $A(y, h, c)$. Here h is determined, given that P will make his maximizing response, and y and c are defined by (2) and (3) as functions of this maximizing h.

To investigate this interactive situation, we must employ some hypothetical utility functions. The selected functions have the partials specified in the discussion of (1), the expression that gives the general form of the utility functions. Two of P's indifference curves derived from his hypothetical utility function are presented in Figure 2.

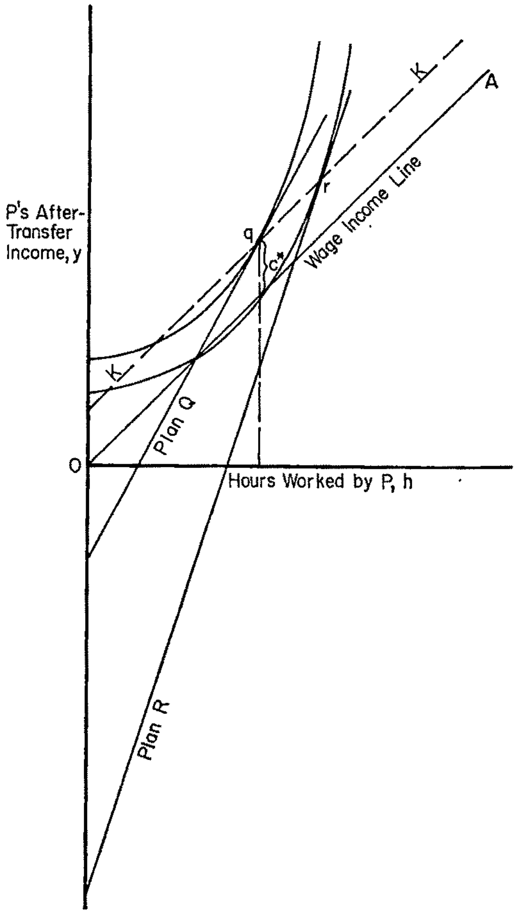


FIGURE 2. TRANSFER PLANS OF EQUAL COST

Each of the possible income transfer plans A could establish can be represented as a straight line across the diagram, which is called a plan line. The base stipend, s, for the plan will give its vertical intercept, the amount P receives if he works not at all. The slope of the plan line indicates the amount of extra income P receives for an additional hour of work. Thus, this slope minus the wage rate gives the marginal tax rate, t, for the plan.

---

tenance plan. This contradicts the hypothesized observation.

If the recipient works more under the income maintenance plan than he would under the NIT, it must be because he is functioning on the CD segment of the income maintenance plan. On this segment the price of leisure is greater than it is with the NIT; the inferior good explanation would not be relevant.

[9] P's reaction function has two arguments; it cannot be represented in a two-dimensional diagram. However, every s, t pair determines a unique plan line. As Figure 2 illustrates, P's reaction to a plan line can be shown in a two-dimensional diagram.

Assume that A established the transfer plan represented as plan Q in the figure. P will choose point q, the point along the plan line that offers P the highest utility, the point where it is tangent to one of his indifference curves.

The cost of this plan to A will be the difference between P's after-transfer income and his wage income. It is shown as c^* on the figure. The line KK shows all combinations of after-transfer income and hours worked for P which cost A the amount c^*. Any plan that leads P to make a maximizing response that lies on KK will cost the same amount. Observe that P's maximizing response to plan R is at point r, a point on KK. This means that plan R will cost A the same amount as plan Q.

Characteristics of the Optimal Plan

To determine the characteristics of A's optimal plan, let him first deal with the single family of plans that cost c^*. Given a choice between transfer plans Q and R, A would decide whether he preferred point q or point r. The two points cost the same. Point r yields P a higher after-transfer income; A values P's income positively. Point r also represents a greater work effort for P; A would likely regard this as an asset as well. With two areas of advantage and none of disadvantage, A would choose plan R over plan Q.

What properties accompany this preference? The base stipend for plan R is significantly more negative than the one for plan Q. The marginal tax rates for both plans are negative. This is illustrated by the fact that the slopes of the plan lines, given by $(1-t)w$, exceed the slope of the wage income line. But the marginal tax rate for the preferred plan, plan R, is substantially more negative.

Surely Q would not be A's optimal plan in this cost family. He could establish plans that would lead P to make responses that lay much farther out along KK. The respective plan lines would be tangent to P's indifference curves running through the relevant points on KK. In general, we would expect the slopes of the intersecting indifference curves to become steeper the farther out we move on KK.[10] This means the tangent plan lines will have steeper slopes (more negative marginal tax rates) and more negative vertical intercepts (more negative base stipends).

Following this argument to its logical extreme, we would expect A's optimal plan in this cost family to have s and t equal to minus infinity. Before these parameters reached the infinity level, however, P would be working so hard that A would value negatively any additional hours worked. The likely result is that A's optimal plan will have s and t take on substantially, but not ridiculously, negative values.[11]

This within-family optimality property will hold no matter what numerical value is assigned to c^*. This implies that the optimum optimorum plan will also have a substantially negative s and t. Otherwise, there would be a plan within its own cost family to which it would be inferior.

---

[10] A movement out along KK gives P more of one valued variable (after-transfer income) and less of another (leisure). We would expect the marginal rate of substitution of the first for the second, as indicated by the slope of P's indifference curve through the relevant point on KK, to be increasing. Even short-run exceptions to this expectation can be ruled out, if 1) P's indifference curves are strictly convex, if 2) leisure is not an inferior good for P, and if 3) all other goods taken together (what is in effect P's after-transfer income) are not inferior for him. Conditions 2) and 3) insure, respectively, that upward and rightward movements across P's indifference map will never lead to indifference curves of decreasing steepness. The three conditions together insure that northeasterly movements will lead to curves of increasing steepness.

[11] In a more extensive version of this paper, I solved A's optimization problem explicitly for the utility functions given in equation (4) below. P's reaction function turns out to be $h=1500-s/2w(1-t)$; he will work less as s or t increases. Given that A's utility function values increases in h positively, for all values of h, we get the extra result that A's optimal transfer plan has s and t equal to minus infinity.

To find his global optimal plan, A must systematically vary the value of c^*, each time finding the optimal plan within the relevant cost family. He then makes a comparison among these plans to find the one that yields him the highest utility. This will be his optimum optimorum plan.

The fact that A's optimal plan has a negative marginal tax rate should not be surprising.[12] We would come to that conclusion by extending the argument relating to the advantages of *NIT* plans over income maintenance plans. A negative t encourages P to earn more on his own. His resulting higher income offers a greater externality to A; A's bonus-granting procedure is similar to the one used by foundations that give matching grants to support universities and at the same time to increase incentives for alumni contributions.

Results with the Optimal Plan—Parallels with Other Strategic Situations

To determine the optimal assistance program to the poor, I utilized the framework of a two-move, non-zero sum game. In this case, A selected a plan that, when it takes account of P's maximizing responses, leads to the best outcome for A. In this context, I showed that a negative income tax plan with fractional marginal tax rates will have an advantage over its limiting form, the income maintenance plan, a plan with a 100 percent marginal tax rate. Plans with negative marginal tax rates were found to be even more attractive from A's standpoint.

Unfortunately, the plan that is optimal for A in my game-theoretic formulation is in no way optimal for P and A together; it is not a Pareto optimum. The representative citizen and the representative poor man can each take actions that convey positive externalities to the other; they are in a situation of reciprocal externalities. The citizen can increase the transfer he gives to the poor man. The poor man can work longer hours and earn more income, thus indirectly increasing the utility of the citizen. But P acts as a self-interested maximizer and A is only interested in those aspects of P's well-being that enter directly into his own utility function. What results, of course, is an outcome inferior to a cooperative solution.

Situations of reciprocal externalities have received much study in the economics literature. Most frequently the point of departure is the outcome with simultaneous independent adjustment. In this context, Otto Davis and Andrew Whinston discuss two firms each of which conveys production externalities to the other, and James Buchanan and Gordon Tullock consider a community whose members can immunize themselves and thus convey an externality to each other by decreasing the likelihood of exposure to a disease. The simultaneous-independent-adjustment outcome, with each party maximizing in his own self-interest, will be inefficient. The equilibrium has the abstract characteristics of the inefficient, but jointly-dominated, outcome of the well-known prisoners' dilemma. (See R. Duncan Luce and Howard Raiffa, pp. 95–102, and Buchanan.)

In the dynamic aspect of their strategic interaction, Stackelberg's model of leadership duopoly (see William Fellner, pp. 98–119) and A. L. Bowley's model of bilateral monopoly closely resemble the model presented here. There are two key parallels among the three models: One player moves first in each of them. He is the first pick his quantity in the duopoly model. He is the seller who starts by setting the price in the bilateral monopoly model. He is the representative citizen who establishes the assistance plan in the income transfer model. Second, the first mover is assumed to know the reaction function of the player who moves second. In moving first he acts strategically and maximizes against the known reaction. The outcome that is achieved in each of these models is dominated by a cooperative solution. Cooperation in these three models would require that their first mover respectively market a smaller quantity, set a lower price, and establish a more generous transfer plan. The second mover would agree not to maximize in a self-interested way against the choice of the first mover. He would market a smaller quantity in the duopoly model, buy a greater quantity in the bilateral monopoly model, and work longer hours in the income transfer model, than he would if he solely considered his own welfare.

---

[12] Jonathan Kesselman also discussed the attractiveness of a negative marginal tax rate for income transfer plans.

III. A Pareto Optimal Transfer Plan

Fortunately, if Pareto optimality is our goal, our analysis has not led us far astray. Plans with substantially negative marginal tax rates can lead to Pareto optima even though P responds as a self-interested maximizer. Except in very perverse cases, plans with positive marginal tax rates will not result in outcomes that are even close to the Pareto frontier.

To see this most clearly, consider hypothetical utility functions for A and for P. They are[13]

$$(4) \quad {}_AU = A(y, h, c) = y^{4/10}h^{1/10}(3000-c)^{5/10},$$

and

$${}_PU = P(y, h) = y^{1/2}(3000-h)^{1/2}$$

To simplify further discussion, I assume that the wage rate, w, is equal to 1.

Figure 3 presents an indifference curve for P and one for A. These curves are derived from the utility functions specified above. From (3), given w, c is uniquely determined once y and h are known. This enables us to represent an indifference curve for A's three-argument utility function in a two-dimensional figure.

If A establishes plan X, P will search along its plan line to find his point of maximum utility. P's maximizing response will be at point x, where the plan line is tangent to his indifference curve.[14] Point x has another important property. It is a point where one of A's indifference curves is tangent to that of P. This means that it is a Pareto optimal point. Thus, plan X leads P to make a response at a Pareto optimal point. It is a Pareto optimal plan.

Plan X has a negative base stipend, $s = -\$5119.37$, which is indicated by the negative intercept of its plan line with the vertical axis. Its negative marginal tax rate, $t = -2.605$, is shown by the fact that the slope of the plan line is greater than the wage rate (in this case 1).

Each Pareto optimal point, each of the points that lies along the contract curve, can be achieved if the linear plan is established tangent to the two indifference curves at their point of tangency. The linear transfer plans that lead to Pareto optimal points, what we might call the Pareto plans, perform a function similar

---

[13] The parameter assigned the value 3,000 in the utility function for A can be increased (decreased) to make cost a less (more) important consideration for A. This particular utility function for P has a unitary elasticity of substitution. Changing the marginal tax rate will affect hours worked only if there is a non-zero stipend. For positive stipends, lowering the marginal tax rate (an effective wage increase) will increase hours worked and conversely.

[14] P's reaction function is easily derived. He maximizes ${}_PU$. For positive ${}_PU$, he can equally well maximize $({}_PU)^2$.

$$\text{(a-1)} \quad ({}_PU)^2 = 3000s - sh + 3000hw(1-t) - h^2w(1-t)$$

Setting the derivative with respect to h equal to zero, we find

$$\text{(a-2)} \quad -s + 3000w(1-t) - 2hw(1-t) = 0$$

Thus,

$$\text{(a-3)} \quad h = 1500 - \frac{s}{2w(1-t)}$$

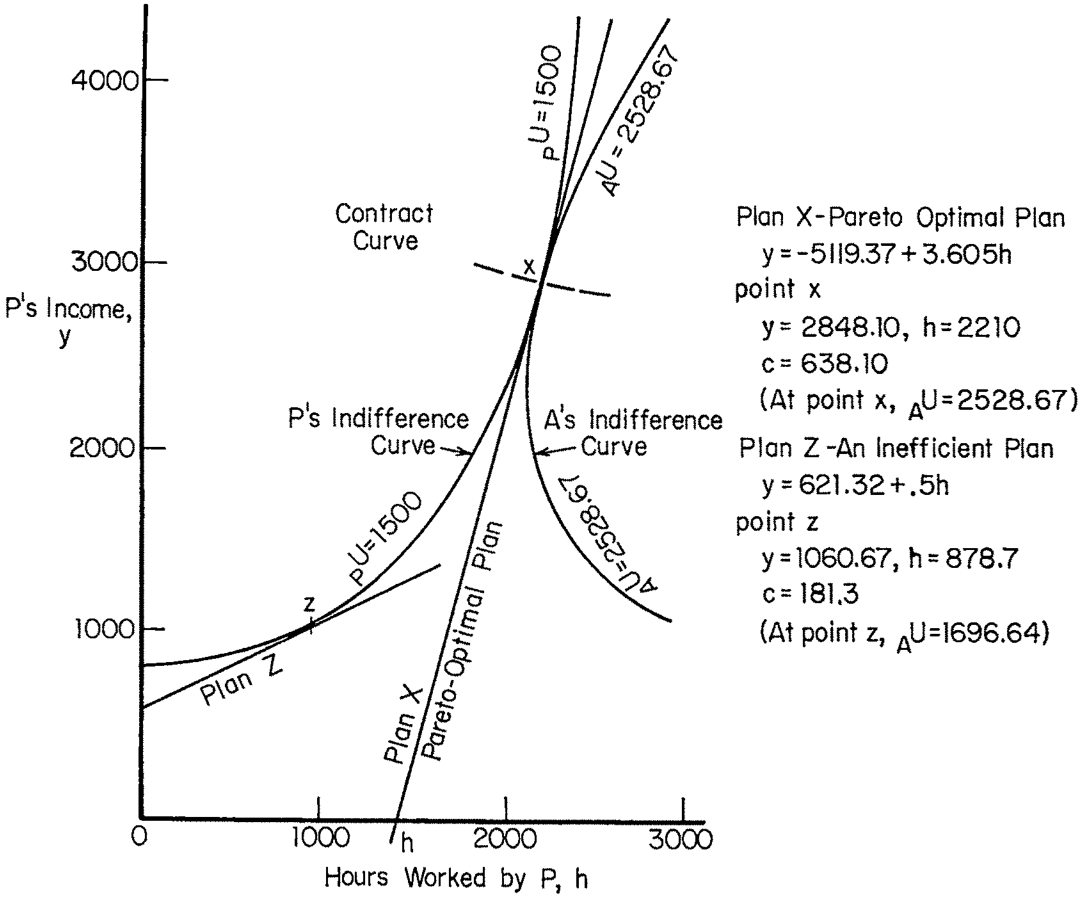


FIGURE 3. ONE PARETO OPTIMAL PLAN

to that of equilibrium price lines in competitive markets. In the case of transfer plans, however, there is no meaningful equivalent to the point of initial endowments.

Of course it would be possible to determine the Pareto optimal points directly, using conventional Lagrangean techniques. Each Pareto optimal point maximizes the sum $P(y, h)+\lambda A(y, h, c)$ for some positive value of λ. Here c is given as function of h and y. For each particular value of λ we can employ conventional maximization techniques to find the y and h that maximize the weighted sum of utilities. By varying the value of λ, we can trace out the locus of efficient trade points, the contract curve.

It is reassuring to observe that simple transfer plans can achieve the same desirable outcomes that could be negotiated directly between A and P. In our society at least, even though the outcomes with the two schemes were the same, it would seem far preferable to establish impersonal pricing schemes that enable individuals to choose how many hours they wish to work, rather than having negotiated outcomes that specify in advance the number of hours to be worked.

Pareto optimal plans have some interesting properties we should note. It is important to understand why there is no need to worry with such plans about the externality that *P*'s work decision conveys to *A*. The function of such plans is to externalize (what James Buchanan and William Stubblebine have labeled) Pareto-relevant externalities. The externalization procedure insures that at the equilibrium point *P* and *A* will have the same trade off rate between the two valued variables, *y* and *h*. This equality in rates guarantees that the equilibrium will be efficient.[15]

The conjunction of three facts implies that linear Pareto optimal plans will have negative marginal tax rates whenever *A*'s and *P*'s utility functions have partials with the signs specified above (whenever their indifference curves have the assumed conventional shape). First, *P*'s indifference curves have positive slope. Therefore, with regard to Pareto optimality considerations, that is with regard to possible points of tangency, we need only consider those portions of *A*'s curves that have positive slope as well. Second, *A*'s indifference curves can never have a section with positive slope that does not exceed the wage rate. Otherwise, northeasterly movements along that section of the curve would indicate more hours worked, more income to *P*, at less cost to *A* (at the same cost if the slope just equals the wage rate). This would represent an improvement for *A* on all three valued variables. It would not be a curve of indifference; *A* would not be indifferent between the old point and the "improved" point. Third, a line that is tangent to an indifference curve of *A* and an indifference curve of *P* at their point of tangency will have the same slope that the indifference curves have at that point. The geometric implication of these three facts is that linear Pareto optimal plans will be represented by lines with positive slopes in excess of the wage rate. With such a plan *P* will be able to work an additional hour and receive extra income in excess of his hourly wage. The marginal tax rate on his earnings will be negative.

Plans with positive marginal tax rates are not only not optimal, they are substantially inferior to the Pareto optimal plans, all of which have negative marginal tax rates. Plan *Z* in Figure 3 is the plan with a marginal tax rate of .5 that gets *P* to the same indifference curve he would reach with plan *X*, a Pareto optimal plan.[16] The base stipend of plan *Z* is $621.32. Under plan *Z*, *P* will choose to work the number of hours and receive the income indicated by point *z*. The geometry makes it evident that for *A*, point *z* is far inferior to point *x*.[17]

Earlier we found that the best plan for the representative citizen will have a negative marginal tax rate. We can conclude from this section that this result still holds if the recipient's welfare, as well as that of the representative citizen, is considered.

IV. The Model in Application

My formulation has overlooked the significant fact that the same employment opportunities are not available to all people. A severely handicapped individual, a member of the hard core unemployed, or a person who receives a very meager wage in accord with his marginal product will derive little benefit from an assistance plan that channels all of its positive transfer through a substantial bonus on wage income.

---

[15] For example, we might encounter an externality-based inefficiency if my neighbor and I disagreed on the desirability of his spending $1,000 annually to keep his house freshly painted. The efficient arrangement would be for me to compensate him at the margin so that his after-compensation trade off rate was the same as mine.

[16] There is no plan with a .5 marginal tax rate which leads to a point on the indifference curve of *A* which runs through point *x*.

[17] A simple calculation reveals the magnitude of the inferiority. If we hold *y*, *h*, and *c* to their values at *z*, it would take a lump sum payment of $3,672 (which *A* could apply against cost) to get *A* to his indifference curve which runs through point *x*. That is, $A(1060.67, 878.7, 3000-(181.3-3672))=2528.67$, the value of $_AU$ at point *x*.

In a more extensive version of this paper, I have shown when the opportunities for poor men differ, the optimal income transfer program will offer more than one transfer plan. For those with little or no opportunity to earn, a plan with a sizable positive stipend will be used to provide assistance. Those who could find productive work will receive assistance under a bonus incentive plan of the type described in this paper.

REFERENCES

A. L. **Bowley**, "On Bilateral Monopoly," *Econ. J.*, 1928, *38*, 651–59.

J. M. Buchanan, "Cooperation and Conflict in Public-Goods Interaction," *Western Econ. J.*, Mar. 1967, *5*, 109–21.

——— **and W. C. Stubblebine**, "Externality," *Economica*, Nov. 1962, *29*, 371–84.

——— **and G. Tullock**, "Public and Private Interaction under Reciprocal Externality," in J. Margolis, ed., *The Public Economy of Urban Communities*, Washington 1965, 52–73.

O. A. Davis and A. Whinston, "Externalities, Welfare, and the Theory of Games," *J. Polit. Econ.*, June 1962, *70*, 241–62.

P. A. Diamond, "Negative Taxes and the Poverty Problem—A Review Article," *Nat. Tax J.*, Sept. 1968, *21*, 288–303.

W. Fellner, *Competition Among the Few: Oligopoly and Similar Market Structures*, New York 1949.

M. Friedman, *Capitalism and Freedom*, Chicago 1962.

C. Green, *Negative Taxes and the Poverty Problem*, Washington 1967.

J. Kesselman, "Labor-Supply Effects of Income, Income-Work, and Wage Subsidies," *J. Hum. Resources*, summer 1969, *4*, 275–92.

R. D. Luce and H. Raiffa, *Games and Decisions*, New York 1957.

M. Olson, Jr., *The Logic of Collective Action: Public Goods and the Theory of Groups*, New York 1968.

P. A. Samuelson, "The Pure Theory of Public Expenditure," *Rev. Econ. Statist.*, Nov. 1954, *36*, 387–89.

R. Theobald, *The Guaranteed Income*, New York 1966.

J. Tobin, "The Case for an Income Guarantee," *Publ. Interest*, summer 1966, No. 4, 31–41.

——— **et al.**, "Is a Negative Income Tax Practical?", *Yale Law J.*, Nov. 1967, *77*, 1–27.

J. Vadakin, "A Critique of the Guaranteed Annual Income," *Publ. Interest*, spring 1968, No. 11, 53–66.

R. Zeckhauser, "Optimal Mechanisms for Income Transfer," RAND P-3878, June 1968, and Harvard Inst. Econ. Res. disc. paper 40, Aug. 1968.

——— **and P. Schuck**, "An Alternative to the Nixon Income Maintenance Plan: A Solution to the Problem of Work Incentives," *Publ. Interest*, spring 1970, No. 19, 120–130.

Poverty Amid Plenty: The American Paradox, *The Report of the President's Commission on Income Maintenance Programs*, Ben Heineman, chairman, Nov. 12, 1969, Washington.

The Optimal Quantity of Money, Bonds, Commodity Inventories, and Capital

By EDGAR L. FEIGE AND MICHAEL PARKIN*

Economists have long concerned themselves with the problem of finding a satisfactory means of incorporating money into the theory of economic behavior. That this problem is still far from being solved is apparent in the surprisingly large discrepancy between the rhetoric of the monetary economist and his analysis. The hallmark of money as extolled in most textbooks is its function as a medium of exchange—yet we have very little in the way of an explicit theory of exchange. Money, that "unique" asset, is generally treated as just another consumer durable good which yields a flow of ill-defined nonobservable services offering the individual "convenience, security and liquidity." The traditional neoclassical approach, underlying much of the important work of several authors (see articles by Milton Friedman, Harry Johnson, Don Patinkin, and Paul Samuelson), introduces real cash balances directly into the utility function, and pays lip service to the costs of exchanges between money and goods although these costs are not made explicit in the analysis. However, once money is treated as an argument of the utility function, many of the interesting questions concerning what money does, and why people hold it—are essentially stifled, since the model does not specify the explicit role money plays as a medium of exchange.

An alternative approach dealing more directly with the transaction aspects of money is the Baumol-Tobin inventory theoretic framework (see articles by William Baumol and James Tobin). The Baumol-Tobin analysis is framed in terms of the explicit costs of making transactions in the money-bond market and thus represents a beginning toward a theory of exchange behavior or its analytic equivalent of optimal inventory theory. Although the theory focuses explicit attention on the costs of exchanging bonds and money, it ignores the costs of exchanging money for commodities and thus neglects the holdings of commodity inventories. Baumol and Tobin leave the utility function unspecified and as such treat total consumption expenditures as exogenously determined. Money is left to enter a budget constraint which involves minimum cost management of bond and money inventories.

Recently, a number of economists have refocused their attention on the rather fundamental problem of formulating a useful theory of the role of money in individual behavior (see articles by Robert Clower, Feige and Donald Nichols, Friedman (1969), Johnson, David Laidler, Alvin Marty, Samuelson). The impetus for this work has been the formulation of a provocative problem in monetary theory; namely, what is the optimal quantity of money and how can individuals be induced to hold the optimal quantity? Furthermore, in what sense is this an optimal quantity of money and what are the consequences of its achievement?

---

* Associate professor of economics, University of Wisconsin; professor of economics, University of Manchester, respectively. The authors wish to acknowledge helpful discussions with Laurits Christensen, David Laidler, Donald Nichols, Kenneth Smith, and the excellent assistance of Yung San Lee and Rodney J. Barrett. The Federal Deposit Insurance Corporation, the University of Wisconsin Graduate School, and the Social Science Research Council (U.K.), provided financial support for the project. An earlier version of this paper was presented in March 1969 at the University of Essex.

This paper has two aims. First, to extend the Baumol-Tobin inventory theoretic approach to incorporate costly transactions between money and commodities. Second, using this broadened inventory theoretic framework, to examine the question of what is the optimal quantity of money.

Our extension of the Baumol-Tobin inventory model denies that money yields any direct services which enter as arguments of the utility function. Money is simply used to facilitate transactions in the bond and commodity markets; and, due to its technological property of being an efficient medium of exchange, its use allows the individual to economize on other real resources which would otherwise be expended in the process of exchange. In this view, the quantity of money enters into the individual's budget constraint, and his inventory management decisions affect the amount of real income available for consumption of other goods and services.

Our aim is to examine this alternative specification of how money affects individual behavior. We wish to determine whether or not the inventory theoretic approach gives rise to conclusions concerning the optimal quantity of money, which are similar to those offered by the neoclassical approach.

It has been claimed by numerous authors that, to the extent that the marginal social cost of increasing the supply of money is zero, individuals should hold money to the point of satiety (see articles by Johnson and Samuelson)—that is, to the point where the marginal utility of holding money is zero. Since money does not bear interest, the opportunity cost to the individual holding part of his wealth in the form of cash balances is the foregone interest on those cash balances. Individuals will thus be induced to economize on cash balance holdings, and the resulting quantity of real cash balances in the society will be less than the Pareto optimal quantity. It is precisely these attempts to economize on real cash balances that give rise to the using up of real resources for transactions purposes and thus to a welfare loss. From society's point of view it is virtually costless to increase the nominal supply of money; but the *real* quantity in existence is in fact determined by the demanders of money. Thus, if the government is to induce people to hold larger real cash balances it must provide an institutional inducement which may or may not be costless. One way in which the community can be persuaded to hold more real cash balances is to pay a rate of interest on money. Alternatively (and equivalent analytically) the government can induce a continuous deflation such that there is an implicit rate of return in money holdings due to the expected increase in their value.

The following analysis concerns itself entirely with the use of paying interest on money as a means of encouraging the community as a whole to hold larger real cash balances. Thus, the "supply of *real* cash balances" can be increased only by affecting the "quantity demanded" of real cash balances, and that is accomplished by paying interest on money. The institutional arrangement which provides for interest payments is unlikely to be costless. If it were costless we would expect the optimal quantity of real cash balances to be the satiety level of real cash balances, i.e., that level of real cash balances which makes the marginal utility of money balances go to zero. This is perfectly consistent with Pareto-efficiency: if the marginal social cost of producing real balances is zero, then the real balances are a free good and marginal utility should be equated with marginal social cost.

The foregoing arguments are generally well known, but there is little evidence to suggest that their implications are well understood. For example, what other consequences flow from real cash balances being a free good? We know that the quantity of real cash balances will increase, but we know little explicitly about the other effects of such a policy. Samuelson has speculated that a satiety policy would result in "fewer trips to the bank and to the broker, . . . smaller printing costs and other costs of transactions" (p. 10). Such explanations, however, are entirely *ex post* rationalizations and certainly cannot be deduced from the current specification of the theory. A major difficulty lies in the manner in which money is typically introduced into the theory of economic behavior. Money is either treated as a consumer durable which yields an unobservable flow of nonpecuniary services or (analytically equivalent and equally unilluminating for the problem at hand) incorporated into the production function as just another input.

Both formulations have contributed something to our general understanding. The consumer theory approach has provided the underlying rationalization for most of the empirical studies of the demand for money—arguing that, as a first approximation, one can regard the flow of nonpecuniary services as proportional to the stock of real cash balances which are observable and thus bring standard consumer theory to bear on the question of the demand for money. When money is included in the utility function for individuals, it can be demonstrated that individuals will hold a less-than-optimal quantity of money. However, the sources of the gains from inducing individuals to hold an optimal quantity of money cannot be directly deduced from such a weakly specified analysis.

Our extension of the inventory theory framework seeks to avoid this problem by explicitly examining the effects of a satiety or free money policy on the quantity of real cash balances held as inventories, as well as its effects on the holdings of bonds, commodity inventories, and productive capital. The analysis also highlights how a free money policy will increase some types of transactions while reducing others. Finally, the theory is capable of yielding exact expressions for the net gain in welfare resulting from a free money policy. The analysis abstracts from problems of risk and uncertainty and also works within the framework of the stationary state.

I. An Extended Inventory Theoretic Model of the Demand for Money, Bonds, and Commodity Inventories—The Individual Optimum

Throughout, we assume that the monetary authority controls the *nominal* money stock in such a way as to maintain a stable price level. The monetary authority is permitted to influence the demand for *real* money balances by paying interest on its monetary obligations. When it does so, or when it varies that rate, it is assumed to simultaneously vary the *nominal* money supply in such a way that the price level is undisturbed. It is worth noting that this assumption is exactly the opposite of that used in much of the growth literature, which treats the rate of inflation (deflation) as a policy parameter and interest rates as market determined. We, instead, treat the rate of interest on the government's monetary obligations as the policy parameter. Thus our analysis abstracts entirely from the dynamics of anticipated and unanticipated changes in the price level, and allows us to treat real cash balances and nominal cash balances interchangeably. A similar correspondence holds between real bond holdings and nominal bond holdings. We could, of course, achieve analytically equivalent results by letting the authorities reduce the nominal money stock at a constant rate, thereby generating a constant rate of (fully anticipated) deflation in the price level.

In addition to assuming fixed prices, we assume a fixed single period of analysis and stationarity.

The economy consists of representative family units which maximize the utility function

$$U = U(pq) \tag{1}$$

where U is the utility index, and pq is the dollar volume of commodities consumed.

The utility function is maximized subject to the constraint

$$0 = Y + \pi - pq - T \tag{2}$$

where

Y = labor income in current dollars
π = net profit from inventory management
pq = expenditure on commodities
T = taxes

Taxes, which are viewed as exogenous by the individual family unit comprise a neutral transfer component and a component involving the real resource costs of operating an interest payments mechanism on balances. Thus

$$T = t + h(r_m), \tag{3}$$

where

t = transfer component
$h(r_m)$ = real resource component
$h(r_m) = 0$ for $r_m = 0$
$h(r_m) = k$ for $r_m > 0$

It is the variable π which represents the novelty of the inventory theoretic approach, and through which we introduce explicitly the role of money as it influences individual behavior. We define π as the profit from inventory management of capital, bonds, money and commodities, net of transaction costs and inventory carrying costs. Specifically,

$$\begin{aligned} \pi = {} & (r_k - \alpha_k)p\overline{K} + (r_b - \alpha_b)\overline{B} \\ & + (r_m - \alpha_m)\overline{M} - \alpha_q\overline{pQ} \\ & - \beta_b n - \beta_q m, \end{aligned} \tag{4}$$

where

$p\overline{K}$ = average stock of capital in current dollars
$\overline{B}$ = average inventory of bonds in current dollars
$\overline{M}$ = average inventory of cash balances
$\overline{pQ}$ = average inventory of commodities in current dollars
r_k = rate of return on capital
r_b = rate of return on bonds
r_m = rate of return on cash balances
α_k = the cost per dollar of carrying capital inventories
α_b = the cost per dollar of carrying bond inventories
α_m = the cost per dollar of carrying cash balances
α_q = the cost per dollar of carrying commodity inventories
n = number of bond market transactions per period
m = number of commodity market transactions per period
β_b = the cost per bond market transaction
β_q = the cost per commodity market transaction.

Most of the variables in the definition of π are familiar; however, the interpretation of the α's and β's require some further elaboration. The α's are inventory carrying cost parameters which reflect the real resource costs of carrying inventories of money, bonds, commodities, and physical capital. They might include such items as the costs of insurance, safety deposit boxes, refrigeration services, and other costs incurred in the storage of inventories. The β's reflect the costs associated with particular transactions such as time, brokerage fees, and delivery charges.

We assume that the individual receives income at the beginning of the period and that he spreads out his total consumption, pq, evenly throughout the period. In order to consume, the individual must maintain some stock of commodity inventories; and we assume that the individual runs down his inventory of commodities at a constant rate. The individual acquires commodity inventories by making m equally spaced trips to the store. Since the total value of consumption for the period is pq, the amount spent per trip to the store is simply pq/m and on average, commodity inventories for the period are

$$p\overline{Q} = \frac{pq}{2m} \tag{5}$$

We assume that all receipts come in the form of money. The individual makes, during the period, n equally spaced trips to the broker. On the first trip he buys bonds. On the subsequent $n-1$ trips he sells bonds. On the first trip he leaves himself with pq/n of cash, buying bonds of $(n-1/n)pq$. On the subsequent $n-1$ trips he cashes pq/n of bonds. Having received pq/n in cash (retained pq/n in the case of the first trip), the individual immediately spends pq/m on commodities. Thus he starts out after a trip to the broker with $pq/n-pq/m$ in cash. Since he runs down this cash with evenly spaced trips to the store, his average cash balance is

$$\overline{M} = \frac{pq}{2n} - \frac{pq}{2m} \tag{6}$$

Finally, average holdings of bond inventories for the period are simply $(n-1/2n)pq$, so that,

$$\overline{B} = \frac{pq}{2} - \frac{pq}{2n} \qquad n \geqq 1 \tag{7}$$

It should be noted that m and n are integers and that m is a multiple of n. The condition that $n=m$ implies that there is perfect synchronization between bond market and commodity market transactions, such that cash balance inventories are zero.

If the broker is regarded as a bank which provides checking services as well as bond market brokerage services, we can give the analysis a somewhat broader interpretation. In this case the individual is assumed to be paid in cash and makes a trip to the banker-broker to deposit cash of pq/n while purchasing bonds of $(n-1/n)pq$. If commodity purchases are paid by check, we consider the marginal cost of checking services to be associated with trips to the store and thus reflected in β_q. In this case β_b is left to reflect the initial cost of establishing a bank account and the brokerage costs of subsequent conversions of bonds into checking deposits. For the special cases where $n=1$, the individual simply deposits his entire income at the banker-broker and purchases no bonds. Thus the average deposit balance inventory becomes $pq/2-pq/2m$ and average bond holdings for the period are zero.

The individual is also subject to a wealth constraint such that his average fixed stock of assets is denoted as

$$\overline{A}^* = p\overline{K} + \overline{B} + \overline{M} + p\overline{Q} \tag{8}$$

Thus,

$$p\overline{K} = \overline{A}^* - \frac{pq}{2} \tag{9}$$

Equation (8) captures the opportunity cost to the individual of adding to cash balances or commodity inventories, because any such transfer can only be accomplished by sacrificing an equivalent dollar volume of bonds or capital.

The representative family must therefore choose the pq, m, and n that will maximize the utility function subject to his budget and wealth constraints. Substituting (4), (5), (6), (7), and (8) into (2), we form the Lagrangean:[1]

$$
\begin{aligned}
V = {} & U(pq) \\
& + \lambda\Bigg[Y + (r_k - \alpha_k)\left(\bar{A}^* - \frac{pq}{2}\right) \\
& + (r_b - \alpha_b)\left(\frac{pq}{2} - \frac{pq}{2n}\right) \\
& + (r_m - \alpha_m)\left(\frac{pq}{2n} - \frac{pq}{2m}\right) \\
& - \alpha_q \frac{pq}{2m} - \beta_b n - \beta_q m - pq - T \Bigg]
\end{aligned}
\tag{10}
$$

Differentiating (10) with respect to pq, n, m, and λ and setting the partial derivatives equal to zero gives the system of equations (11).

In order to compare the results implied by equations (11) with the more familiar Baumol inventory model, we first examine a partial solution which considers pq as predetermined. Under this assumption, equations (11b) and (11c) can be solved for n and m as functions of pq to determine the optimal number of trips to the broker and the store:

$$
n^* = \sqrt{\frac{(r_b - \alpha_b - r_m + \alpha_m)}{2\beta_b}\, pq} \tag{12}
$$

$$
m^* = \sqrt{\frac{(r_m - \alpha_m + \alpha_q)}{2\beta_q}\, pq} \tag{13}
$$

Assuming a given level of consumption, equations (12) and (13) reveal how interest payments on cash balances affect transactions behavior. Non-zero interest payments on money will reduce trips to the broker but will increase trips to the store. Conversely, as the costs of carrying money balances increase, a greater number of

$$
(11)\quad
\begin{aligned}
&\text{(a)}\quad V_{pq} = U'(pq) - \lambda\left[\frac{(r_k - \alpha_k - r_b + \alpha_b)}{2} + \frac{(r_b - \alpha_b - r_m + \alpha_m)}{2n}\right. \\
&\qquad\qquad \left. + \frac{(r_m - \alpha_m + \alpha_q)}{2m} + 1\right] = 0 \\
&\text{(b)}\quad V_n = \lambda\left[\frac{(r_b - \alpha_b - r_m + \alpha_m)}{2n^2}\, pq - \beta_b\right] = 0 \\
&\text{(c)}\quad V_m = \lambda\left[\frac{(r_m - \alpha_m + \alpha_q)}{2m^2}\, pq - \beta_q\right] = 0 \\
&\text{(d)}\quad V_\lambda = \left[Y + (r_k - \alpha_k)\left(\bar{A}^* - \frac{pq}{2}\right) + (r_b - \alpha_b)\left(\frac{pq}{2} - \frac{pq}{2n}\right)\right. \\
&\qquad\qquad \left. + (r_m - \alpha_m)\left(\frac{pq}{2n} - \frac{pq}{2m}\right) - \alpha_q\frac{pq}{2m} - \beta_b n - \beta_q m - pq - T\right] = 0
\end{aligned}
$$

---

[1] Although inventory theory specifies n and m as integers, we treat the Lagrangean as continuously differentiable. This treatment can result in some minor departures from the minimum cost conditions. We chose to ignore this difficulty and throughout the paper allow n and m to take on non-integer values as determined by the parameters of the model.

bond market transactions will be undertaken and a smaller number of commodity market transactions will be made. Thirdly, increased transactions costs will reduce the number of transactions in the respective markets.

Substituting (12) and (13) into (5), (6), and (7) yields partial demand functions for average cash balances, average commodity inventories and average bond holdings. These are:

$$(14)\quad (\overline{M})^* = \sqrt{\frac{\beta_b pq}{2(r_b - \alpha_b - r_m + \alpha_m)}} - \sqrt{\frac{\beta_q pq}{2(r_m - \alpha_m + \alpha_q)}}$$

$$(15)\quad (p\overline{Q})^* = \sqrt{\frac{\beta_q pq}{2(r_m - \alpha_m + \alpha_q)}}$$

$$(16)\quad (\overline{B})^* = \frac{pq}{2} - \sqrt{\frac{\beta_b pq}{2(r_b - \alpha_b - r_m + \alpha_m)}}$$

Equations (12) to (16) suggest that a change in r_b directly affects the number of trips to the broker but does not affect the number of trips to the store. Hence, bond and money inventories are switched but commodity inventories are unaffected.[2] That is, a change in r_b directly affects only the margin of substitution between bonds and money. A change in r_m affects the margin of substitution between money and bonds as well as between money and commodity inventories.

Consider the case where both r_b and r_m rise by the same amount leaving invariant $(r_b - r_m)$. Traditionally, $(r_b - r_m)$ has been treated as the opportunity cost of holding real balances; an unchanged $(r_b - r_m)$ would leave invariant the stock of real balances held. This result remains intact in the Baumol model even when explicit interest is paid on money. But when the model is extended to the margin of substitution between money and commodities, an increase in r_b and r_m which leaves the difference unchanged, alters the stock of real balances held. This result arises from the fact that r_m affects the margin of substitution between bonds and money on the one hand, and between money and commodities on the other, whereas r_b affects only the margin between bonds and money. This implies that the demand function for real balances is not a unique function of the difference between r_b and r_m but instead depends on their levels. Of course, what has to be noticed here is that in an n asset world, there are $n-1$ margins of substitution and only if *all* rates move by the same absolute amount will demands be invariant to the rate levels. In our model, we can interpret α_q (the carrying cost of commodity inventories) as the (negative) rate of return on commodity inventories. An increase in r_b and r_m with an equal change (decrease) in α_q will in fact leave all quantities demanded unchanged.

The demand function for cash balances (14) reduces to the familiar Baumol result given Baumol's more restrictive assumptions.

In Baumol's model, pq is predetermined; there are no costs associated with the purchase of commodities; no interest payments on real cash balances; and no costs of carrying money or bond inventories. Thus, $\beta_q = r_m = \alpha_m = \alpha_b = 0$, and (14) reduces to the Baumol demand function,

$$(17)\quad (\overline{M})^* = \sqrt{\frac{\beta_b pq}{2r_b}} = \frac{1}{2}\sqrt{\frac{2\beta_b pq}{r_b}}$$

In order to determine the individual's final equilibrium values of n, m, and pq, we drop the assumption that pq is predetermined, and simultaneously solve the first-order conditions of equation (11) subject to two additional constraints that must be satisfied for the society as a whole. The first constraint simply states that taxes must equal the interest payments on bonds and money plus the real resource costs of instituting an interest payments mechanism on money. Thus:

---

[2] Note this is only true when pq is predetermined. When pq is endogeneously determined, a change in r_b will have an indirect effect on commodity inventories due to the induced change in total income.

$$T = r_b\overline{B} + r_m\overline{M} + h(r_m), \tag{18}$$

where

$$h(r_m) = 0 \qquad \text{for } r_m = 0$$
$$h(r_m) = k \qquad \text{for } r_m > 0$$

and the tax transfer scheme is assumed to have neutral allocation consequences.

The final constraint derived from the stationary assumption, states that society's stock of inventories of physical capital plus the stock of commodity inventories is constant.[3] Thus:

$$\overline{W}^* = p\overline{K} + p\overline{Q} \tag{19}$$

or

$$\overline{pK} = \overline{W}^* - \frac{pq}{2m} \tag{19'}$$

$$(pq)^* = \frac{X \pm \sqrt{X^2 - 4A^2B^2}}{2B^2} \tag{20}$$

where

$$\begin{aligned} X = {} & (2 + \alpha_b)[Y + (r_k - \alpha_k)\overline{W}^* - h(r_m)] \\ & + \left[\frac{(\alpha_m - \alpha_b)^2}{2(r_b - \alpha_b - r_m + \alpha_m)} + \frac{1}{2}(r_b - r_m) + \frac{3}{2}(\alpha_m - \alpha_b)\right]\beta_b \\ & + \left[\frac{(r_k - \alpha_k - \alpha_m + \alpha_q)^2}{2(r_m - \alpha_m + \alpha_q)} + \frac{1}{2}(2r_k - 2\alpha_k + r_m - 3\alpha_m + 3\alpha_q)\right]\beta_q \\ & + \left[\sqrt{\frac{(\alpha_m - \alpha_b)^2(r_k - \alpha_k - \alpha_m + \alpha_q)^2}{(r_m - \alpha_m + \alpha_q)(r_b - \alpha_b - r_m + \alpha_m)}} + \sqrt{\frac{(\alpha_m - \alpha_b)^2(r_m - \alpha_m + \alpha_q)}{(r_b - \alpha_b - r_m + \alpha_m)}}\right. \\ & + \sqrt{\frac{(r_k - \alpha_k - \alpha_m + \alpha_q)^2(r_b - \alpha_b - r_m + \alpha_m)}{(r_m - \alpha_m + \alpha_q)}} \\ & \left. + \sqrt{(r_m - \alpha_m + \alpha_q)(r_b - \alpha_b - r_m + \alpha_m)}\right]\sqrt{\beta_b\beta_q} \end{aligned}$$

$$A = Y + (r_k - \alpha_k)\overline{W}^* - h(r_m)$$

$$B = \tfrac{1}{2}(2 + \alpha_b)$$

$$n^* = \sqrt{\frac{(r_b - \alpha_b - r_m + \alpha_m)}{2\beta_b}\, pq,^*} \tag{21}$$

$$m^* = \sqrt{\frac{(r_m - \alpha_m + \alpha_q)}{2\beta_q}\, pq^*} \tag{22}$$

---

[3] It should be noted that $\overline{W}^*$ is not the same thing as accounting net worth. If $r_m=0$, accounting net worth is $\overline{W}^*+\overline{M}$. With $0<r_m<r_k$, $\overline{W}^*<$ accounting net worth $<\overline{W}^*+\overline{M}$.

Simultaneous solution of (11), (18), and (19) yields quadratic solution equations for determining the equilibrium values of pq, n, and m.[4] These are shown in equations (20)–(22).

Substituting equation (20) into equations (14), (15), and (16) reveals that the demand for average cash balances, commodity inventories, and bonds, depends upon disposable human and nonhuman income—as well as interest rates, transactions costs, and inventory carrying costs.

II. The Social Optimum

The foregoing analysis has determined the optimal equilibrium values of consumption, number of transactions, and inventory holdings, for the individuals in society. In order to determine the corresponding optimal values for society as a whole, it is necessary to examine the sources and uses of resources for the entire society. Given the stationarity constraint and assuming a total population of N families, the total nonhuman wealth for the society is simply:

$$N\overline{W}^* = N(\overline{pK} + \overline{pQ}), \tag{23}$$

and the total income available to society is

$$N[Y + r_k(\overline{W}^* - \overline{pQ})] \tag{24}$$

Thus, for society as a whole, the opportunity cost of holding commodity inventories is the income foregone by not utilizing these inventories as productive capital. The society must allocate its resource flow among the following uses:

(a) holding money at a cost of

$$N\left[\alpha_m\left(\frac{pq}{2n} - \frac{pq}{2m}\right)\right]$$

(b) holding commodity inventories at a cost of

$$N\left[\alpha_q\left(\frac{pq}{2m}\right)\right]$$

(c) holding bond inventories at a cost of

$$N\left[\alpha_b\left(\frac{pq}{2} - \frac{pq}{2n}\right)\right]$$

(d) holding capital inventories at a cost of

$$N\left[\alpha_k\left(W^* - \frac{pq}{2m}\right)\right]$$

(e) transacting in bond markets at a cost of $N\ [\beta_b n]$

(f) transacting in commodity markets at a cost of $N[\beta_q m]$

(g) operating a scheme of paying interest on money of $N[h(r_m)]$, and

(h) consuming, $N[pq]$

The social constraint therefore becomes:

$$\begin{aligned} 0 = N\Bigg[& Y + r_k\left(\overline{W}^* - \frac{pq}{2m}\right) \\ & - \alpha_m\left(\frac{pq}{2n} - \frac{pq}{2m}\right) \\ & - \alpha_q\left(\frac{pq}{2m}\right) - \alpha_b\left(\frac{pq}{2} - \frac{pq}{2n}\right) \\ & - \alpha_k\left(\overline{W}^* - \frac{pq}{2m}\right) \\ & - \beta_b n - \beta_q m - h(r_m) - pq\Bigg] \end{aligned} \tag{25}$$

The solution to the social problem is obtained by maximizing the Lagrangean:

---

[4] It will be noticed that we can solve $(pq)^*$ despite the fact that an unspecified function of $pq - U'(pq)$ appears in the first-order conditions. This is because the first-order conditions form a recursive system and (11a) solves for λ (the Lagrange multiplier) given the values of the other variables from (11b), (11c), and (11d).

$$(26)\quad V = N\Bigg\{U(pq) + \lambda\Bigg[Y + (r_k - \alpha_k) \cdot\left(\overline{W}^* - \frac{pq}{2m}\right) - \alpha_m\left(\frac{pq}{2n} - \frac{pq}{2m}\right) - \alpha_b\left(\frac{pq}{2} - \frac{pq}{2n}\right) - \alpha_q\frac{pq}{2m} - \beta_b n - \beta_q m - h(r_m) - pq\Bigg]\Bigg\}$$

Differentiating V and setting the partial derivatives equal to zero yields the system (27).

The first-order conditions (27) can be simultaneously solved to determine the socially optimum level of consumption expenditures, as well as the socially optimum number of transactions. The solution equations are as follows:

$$(28)\quad Npq^* = N\left[\frac{Z \pm \sqrt{Z^2 - 4A^2B^2}}{2B^2}\right],$$

where

$$Z = [Y + (r_k - \alpha_k)\overline{W}^* - h(r_m)] \cdot(2 + \alpha_b) + 2(\alpha_m - \alpha_b)\beta_b + 2(r_k - \alpha_k - \alpha_m + \alpha_q)\beta_q + 4\sqrt{(\alpha_m - \alpha_b)(r_k - \alpha_k - \alpha_m + \alpha_q)} \cdot\sqrt{\beta_b\beta_q}$$

$$A = Y + (r_k - \alpha_k)\overline{W}^* - h(r_m)$$

$$B = \frac{(2 + \alpha_b)}{2},$$

$$(29)\quad Nn^* = N\sqrt{\frac{(\alpha_m - \alpha_b)}{2\beta_b}\,pq},^*$$

and

$$(30)\quad Nm^* = N\sqrt{\frac{(r_k - \alpha_k - \alpha_m + \alpha_q)}{2\beta_q}\,pq^*}$$

In order to compare the socially optimum number of transactions and inventories with the individual optimum values, we construct Table 1, which summarizes the solutions for the social and individual optimum values as functions of pq.

Table 1 reveals that the only way that individuals can be induced to undertake the socially optimum number of transactions, and (equivalent analytically) hold the socially optimum stocks of inventories, is to pay interest on cash balances and bonds equal to the *net* rate of return on capital. Similarly, comparing equations (20) and (28) reveals that optimum individual consumption will equal optimum social consumption if, and only if,

$$(31)\qquad r_m = r_b = r_k - \alpha_k$$

This result caused us some surprise at first. Why is not the optimality condition

$$(31\text{a})\quad r_m - \alpha_m = r_b - \alpha_b = r_k - \alpha_k?$$

$$(27)\quad \begin{aligned}
&\text{(a)}\quad V_{pq} = N\left\{U'(pq) - \lambda\left[\frac{\alpha_b}{2} + \frac{(r_k - \alpha_k - \alpha_m + \alpha_q)}{2m} + \frac{(\alpha_m - \alpha_b)}{2n} + 1\right]\right\} = 0\\
&\text{(b)}\quad V_n = N\lambda\left[\frac{(\alpha_m - \alpha_b)}{2n^2}pq - \beta_b\right] = 0\\
&\text{(c)}\quad V_m = N\lambda\left[\frac{(r_k - \alpha_k - \alpha_m + \alpha_q)}{2m^2}pq - \beta_q\right] = 0\\
&\text{(d)}\quad V_\lambda = N\Bigg[Y + (r_k - \alpha_k)\left(\overline{W}^* - \frac{pq}{2m}\right) - \alpha_m\left(\frac{pq}{2n} - \frac{pq}{2m}\right)\\
&\qquad\qquad - \alpha_b\left(\frac{pq}{2} - \frac{pq}{2n}\right) - \alpha_q\frac{pq}{2m} - \beta_b n - \beta_q m - h(r_m) - pq\Bigg] = 0
\end{aligned}$$

TABLE 1—OPTIMAL TRANSACTIONS AND INVENTORY HOLDINGS

Variable	Individual Optimum	Social Optimum
$(n)^*$	$\sqrt{\frac{(r_b - \alpha_b - r_m + \alpha_m)pq}{2\beta_b}}$	$\sqrt{\frac{(\alpha_m - \alpha_b)pq}{2\beta_b}}$
$(m)^*$	$\sqrt{\frac{(r_m - \alpha_m + \alpha_q)pq}{2\beta_q}}$	$\sqrt{\frac{(r_k - \alpha_k - \alpha_m + \alpha_q)pq}{2\beta_q}}$
$(\overline{M})^*$	$\sqrt{\frac{\beta_b pq}{2(r_b - \alpha_b - r_m + \alpha_m)}} - \sqrt{\frac{\beta_q pq}{2(r_m - \alpha_m + \alpha_q)}}$	$\sqrt{\frac{\beta_b pq}{2(\alpha_m - \alpha_b)}} - \sqrt{\frac{\beta_q pq}{2(r_k - \alpha_k - \alpha_m + \alpha_q)}}$
$(\overline{B})^*$	$\frac{pq}{2} - \sqrt{\frac{\beta_b pq}{2(r_b - \alpha_b - r_m + \alpha_m)}}$	$\frac{pq}{2} - \sqrt{\frac{\beta_b pq}{2(\alpha_m - \alpha_b)}}$
$(\overline{pQ})^*$	$\sqrt{\frac{\beta_q pq}{2(r_m - \alpha_m + \alpha_q)}}$	$\sqrt{\frac{\beta_q pq}{2(r_k - \alpha_k - \alpha_m + \alpha_q}}$
$(\overline{pK})^*$	$\overline{W}^* - \sqrt{\frac{\beta_q pq}{2(r_m - \alpha_m + \alpha_q)}}$	$\overline{W}^* - \sqrt{\frac{\beta_q pq}{2(r_k - \alpha_k - \alpha_m + \alpha_q)}}$

If r_m and r_b were fixed to achieve the equalities specified in (31a) instead of those in (31), the private opportunity cost of the services from real balances of money and bonds would be zero. But α_m and α_b represent real social opportunity costs of holding money and bonds (and hence of receiving the service yield from those holdings). Optimality requires that the private opportunity costs must equal the social costs, hence the private cost of holding money and bonds must be α_m and α_b, respectively. This is achieved by paying interest at r_m and r_b equal to the net return on physical capital and having the individual bear the carrying costs α_m and α_b.

It may still be objected that (31a) must be satisfied because net rates of return must be equalized. In fact, net rates of return are equalized in (31). They are unequal in (31a). The reason for this is that money and bonds both yield returns in the form of the increased consumption made possible by lowering inventory management costs. These returns are additional to the interest income that will arise from holding money and bonds. The optimality conditions (31) suggest that these additional service yields from money and bonds are equal on the margin to α_m and α_b, respectively.

III. Welfare Implications

We can now explicitly determine the sources and possible magnitudes of the net welfare gain resulting from the institution of a payments mechanism on cash balances which satisfies equation (31). To this end we consider two regimes.

In regime I, we assume that bonds yield the net rate of return on capital, but no interest is paid on cash balances. Regime II is similar to regime I except that interest is paid on cash balances equal to the return on bonds and the net return on capital. We assume for the moment that it is costless to introduce an interest-payments mechanism on money. We define the gross welfare gain to society from the institution of a payments mechanism on money as that increase in consumption which can be indefinitely maintained by shifting from regime I to regime II. The consumption level under the two regimes can be obtained directly by solving the quadratic solution equation (20) under the respective sets of assumptions. To use a numerical illustration, we assume the following values for the parameters of the model:

(32)

$$
\begin{aligned}
Y &= \$10{,}000.00 & \alpha_m &= .01\\
W^* &= \$50{,}000.00 & \beta_b &= \$15.00\\
r_k &= .15 & \beta_q &= .20\\
\alpha_k &= .09 & \alpha_q &= .06\\
r_b &= .06 & h(r_m) &= 0\\
\alpha_b &= .005 & r_m &= 0 \text{ Regime I}\\
N &= 50{,}000{,}000 & r_m &= .06 \text{ Regime II}
\end{aligned}
$$

Solving (20), given the foregoing values of the parameters, we derive two roots for each regime. The higher root exceeds the budget constraint and therefore only the legitimate lower root is presented in Table 2.

TABLE 2—OPTIMUM CONSUMPTION LEVELS

	Per family consumption (dollars)	Total consumption (billion dollars)
Regime I	$12,857.50	$707.163
Regime II	12,899.90	709.495
Maximum Net Gain	42.40	2.332

Given the hypothetical values of the parameters, the maximum gain from paying interest on cash balances is approximately $2.3 billion per year. From this gain must be subtracted the real resource costs of instituting a payments mechanism.[5] In the above example, if these costs should exceed $2.3 billion per year, the society would, of course, be better off if no interest were paid on cash balances.

Given the optimal level of consumption in the two regimes, one can calculate the optimal transactions and optimal inventory stocks in the two regimes. Table 3 presents the figures based on the foregoing assumptions concerning the parameter values.

The effects of paying interest on money become apparent from Table 3. Brokerage transactions are substantially reduced while a larger number of commodity market transactions are undertaken. The supply of cash balances must be increased by over 400 percent with a corresponding decrease in holdings of bonds of about 60 percent. The increase in consumption from regime I to regime II results in part from the fact that nonproductive commodity inventories are reduced and shifted into productive capital. Further gains are achieved through the other components of the net profits from overall inventory management. In order to make the sources of these gains explicit, we present (in Table 4) the total income and expenditure flows under the two regimes.

It is clear from Table 4 that the welfare gain arising from payment of interest on cash balances does not simply come from a higher income on a larger stock of productive capital. The example cited suggests that gains also arise from a reduction in the costs of bond market transactions which outweight both increased commodity market transactions costs and the net increase in overall inventory management costs.

IV. Further Consequences of the Analysis

The suggestion of paying interest on cash balances has led some economists to speculate that individuals would shift their entire financial portfolio out of bonds and into money. Let us, therefore, look at the necessary conditions for bond holdings

---

[5] In his recent contribution on the optimal quantity of money, Friedman (1969) using a completely different analytic approach, estimated the gross welfare gain from paying interest on money to be between $2 and $4 billion per year, assuming a net rate of return on capital of 5 percent. Our own estimates (which are essentially illustrative) suggest a similar gain based on a 6 percent net return on capital. Friedman goes on to estimate the gain using a 17 percent rate of return, and finds the gain to range between $60–$100 billion per year. Our own calculations suggest that at a 15 percent rate of return the gain would be approximately $5–$6 billion per year.

TABLE 3—OPTIMAL TRANSACTIONS AND INVENTORY HOLDINGS

Regime I: $(r_m=0)$
Regime II: $(r_m=r_b=r_k-\alpha_k)$

	$(n)^*$ (units)	$(m)^*$ (units)	$(\overline{M})^*$ (dollars)	$(\overline{B})^*$ (dollars)	$(p\overline{Q})^*$ (dollars)	$(p\overline{K})^*$ (dollars)
Per Family						
Regime I	5.28	40.09	952.16	5316.24	160.35	49839.65
Regime II	1.47	59.56	4290.50	2051.16	108.29	49891.71
	(bil. un.)	(bil. un.)	(bil. dol.)	(bil. dol.)	(bil. dol.)	(bil. dol.)
Totals						
Regime I	.290	2.205	52.37	292.39	8.82	2741.20
Regime II	.081	3.276	235.98	112.81	5.96	2744.04

TABLE 4—REVENUES AND EXPENDITURES
(dollars per family)

Source	Regime I		Regime II		Net Change	
Revenues						
Human Income (Y)	$10,000.00		$10,000.00		$ .00	
Income from capital $r_k(p\overline{K})^*$	7,475.95		7,483.76		+ 7.81	
Income from bonds $r_b(\overline{B})^*$	318.97		123.07		−195.90	
Income from money $r_m(\overline{M})^*$	0.00		257.43		+257.43	
Total Revenues		$17,794.92		$17,864.26		+$69.34
Expenditures						
Inventory Management Costs						
$\alpha_k(p\overline{K})^*$	$ 4,485.57		$ 4,490.25		+$ 4.68	
$\alpha_b(\overline{B})^*$	26.58		10.26		− 16.32	
$\alpha_q(p\overline{Q})^*$	9.62		6.50		− 3.12	
$\alpha_m(\overline{M})^*$	9.52		42.91		+ 33.39	
Subtotal	$ 4,531.29		$ 4,549.92		+$18.69	
Transaction Costs						
$\beta_b n$	$ 79.18		$ 21.99		−$57.19	
$\beta_q m$	8.02		11.91		+ 3.89	
Subtotal	$ 87.20		$ 33.90		+$53.30	
Taxes						
Transfer Taxes						
$r_b(\overline{B})^*+r_m(\overline{M})^*$	$ 318.97		$ 380.50		+$61.53	
Cost of Payment Mechanism $h(r_m)$	$ 0.00		$ 0.00		0.00	
Subtotal	$ 318.97		$ 380.50		+$61.53	
Total Expenditures		$ 4,937.46		$ 4,964.32		+$26.26
Net Profit Available for Consumption Expenditures		$12,857.46		$12,899.94		+$42.48

to be positive when interest is paid on money. Given the definition of bond balances,

$$\overline{B} = \frac{pq}{2} - \frac{pq}{2n}, \tag{33}$$

average bond holdings will be positive so long as $n > 1$. When the equilibrium value is substituted for n from equation (29), the condition becomes

$$n^* = \sqrt{\frac{(\alpha_m - \alpha_b)pq}{2\beta_b}} > 1, \tag{34}$$

which can be written as

$$\alpha_m > \alpha_b + \frac{\beta_b}{pq/2} \tag{35}$$

Thus bond holdings will coexist with money holdings even when interest is paid on cash balances—so long as the cost per dollar of money inventories exceeds the cost per dollar of bond inventories plus the cost per dollar of bond transactions.

A similar analysis reveals the conditions required for money balances to be positive in a stationary economy.[6] Since

$$\overline{M} = \frac{pq}{2n} - \frac{pq}{2m}, \tag{36}$$

money balances will be positive so long as $m > n$. Zero money balances will occur when $m = n$, i.e., when there is perfect synchronization between bond market and commodity market transactions. Money balances will be positive so long as

$$\sqrt{\frac{(r_m - \alpha_m + \alpha_q)}{2\beta_q}\, pq} > \sqrt{\frac{(r_b - \alpha_b) - (r_m - \alpha_m)}{2\beta_b}\, pq}, \tag{37}$$

or

$$\frac{(r_m - \alpha_m + \alpha_q)}{(r_b - \alpha_b) - (r_m - \alpha_m)} > \frac{\beta_q}{\beta_b} \tag{38}$$

Equation (38) states that cash balances will be positive so long as the ratio of net profit from holding money rather than commodities, to net profit from holding bonds rather than money, exceeds the ratio of commodity market to bond market transaction costs.

Finally, it has been suggested that if interest payments are made on money, individuals will cease to spend money balances. The foregoing analysis suggests that interest payments on cash balances will induce individuals to hold larger average cash balances, but that they will spend them more frequently by undertaking a larger number of transactions in commodity markets.

V. Summary and Conclusions

The analysis presented above suggests that Pareto-efficiency requires that interest be paid on cash balances which is equal to the rate of return on bonds and the net rate of return on capital. This conclusion, which has also been derived from the traditional utility maximizing model which includes cash balances in the utility function, emerges unscathed when the problem of determining the optimal quantity of money is subjected to an extended inventory theoretic analysis. The inventory theory approach makes clear that the problem of an optimal money supply is inextricably linked to the problem of optimal bond inventories, commodity inventories, and capital stock.

The inventory theory specification of the optimal money problem is extremely rich. It yields explicit demand functions for money, bonds, commodity inventories, and capital, as well as expressions for the optimal number of transactions in different markets.

Moreover, the inventory theory approach enables one to examine the explicit sources and magnitudes of the welfare gain resulting from an optimal money policy. The major conclusions concerning the effects of paying interest on money may be summarized as follows:

---

[6] B. Pesek and T. Saving have claimed that when interest is paid on money, real cash balances go to zero.

(a) cash balances will increase,
(b) bond holdings will decrease,
(c) commodity inventory holdings will decrease,
(d) physical reproductive capital will increase,
(e) commodity market transactions will increase,
(f) bond market transactions will decrease,
(g) consumption will increase.

These conclusions are, of course, subject to the qualification that the real resource costs of instituting an interest payment mechanism are lower than the potential gross welfare gain from increased consumption. The analysis clarifies the distinction between a) the virtually costless process of increasing nominal balances and b) the potentially costly process of instituting a payment mechanism on money sufficient to induce individuals to hold the socially optimal collection of asset inventories. One obvious mechanism for paying interest on cash balances would be to induce a steady deflation. Such a policy, however, would be very costly both in terms of transitional adjustments as well as resource costs of continuously changing price tags on commodities. A "second best" solution likely to be less costly in terms of real resources and yet be capable of capturing a substantial portion of the potential welfare gain would be to eliminate the present prohibition of interest payments on demand deposits and to allow explicit payment of interest on bank reserves and vault cash. A third solution would be for banks to be permitted to issue their own competitively produced notes.

The inventory theory approach to the optimal money problem as developed in this paper draws explicit attention to the capital costs of inventory management as well as the capital costs of transacting in bond and commodity markets. It still remains to generalize the model to take explicit account of the human income costs of transactions, by investigating effects of an optimal money policy on the work-leisure margin. Finally the model should be generalized to introduce risk and uncertainty so as to capture the precautionary and speculative motives for holding cash balances.

REFERENCES

W. J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Quart. J. Econ.*, Nov. 1952, *66*, 545–56.

R. Clower, "What Traditional Money Theory Really Wasn't," *Can. J. Econ.*, May 1969, *2*, 299–302.

E. Feige and D. Nichols, "Money, Wealth and Welfare," *Social Systems Research Institute Workshop Series*, Firm and Market No. 6828, Univ. Wisconsin 1968.

M. Friedman, *A Program for Monetary Stability*, New York 1960.

———, *The Optimum Quantity of Money*, Chicago 1969.

H. Johnson, "Inside Money, Outside Money, Income, Wealth and Welfare in Monetary Theory," *J. Mon. Cred. Bank*, Feb. 1969, *1*, 30–45.

D. E. W. Laidler, "Money, Wealth and Time Preference in a Stationary Economy," *Can. J. Econ.*, Nov. 1969, *2*, 526–35.

A. Marty, "Inside Money, Outside Money and the Wealth Effect: A Review Essay," *J. Mon. Cred. Bank.*, Feb. 1969, *1*, 101–11.

D. Patinkin, *Money, Interest and Prices*, New York 1965.

B. Pesek and T. Saving, *Money, Wealth and Economic Theory*, New York 1967.

P. A. Samuelson, "What Classical and Neo-Classical Monetary Theory Really Was," *Can. J. Econ.*, Feb. 1968, *1*, 1–15.

J. Tobin, "The Interest Elasticity of Transactions Demand for Cash," *Rev. Econ. Statist.*, Aug. 1956, *38*, 241–47.

A Utility Theory of Representative Government

By Edwin T. Haefele*

The devolution of economic and political meaning to the individual—hence the identification of the individual *act* as the source of economic and political values—culminated with eighteenth century rationalism. It was then that the foundation was laid for an economic theory which was to dominate the Western world. The political theory of representative government, codified by the American Constitution, was formulated at the same time. Both theories were devised by men who, as Lawrence Frank puts it, "... were persuaded by the 18th century belief in the rationality of man and accepted proposals that emphasized the individual's capacity for acting rationally in pursuing his own self-interest and happiness; calculating his prospective gains and losses" (p. 809).

Since that time, the economic theory based on personal utility calculations has prospered while the political theory based on these same calculations has languished and is now suspect. This is not to say that everywhere men still trust a *laissez faire* market economy while mistrusting representative government. It is to say that economic theory, using the concept of a competitive market, can explain much about economic systems both competitive and otherwise, while Anglo-American political theory, having developed no such counterpart concept, has lost its organizing principle.

The loss came very early, for in one lifetime came the brilliant exposition of utility-based political theory, the Federalist Papers; the not-so-"felicific calculus" of Jeremy Bentham; and the fatal utilitarianism of J. S. Mill. Utility-based political theory died aborning, yet the government designed squarely on its precepts has prospered for nearly two hundred years. Prospered because of its utility foundation? Prospered in spite of its utility foundation? Do we know?

In 1835, Alexis de Tocqueville was clear it was because of its utility foundation. "If ... you do not succeed in connecting the notions of right with that of personal interest, which is the only immutable point in the human heart, what means will you have of governing the world except by fear?" (pp. 147–48). The authors of the Federalist Papers were likewise convinced. Lately the question has shifted considerably and split into two parts: does representative government really *have* a utility base, and is a utility base appropriate for the problems of the present? Both questions have important practical as well as theoretical significance and it may prove useful to review why that is so.

Aggregate measures, whether they be *GNP*, personal income, net social return on investment, or whatever, are increasingly disputed because of the distribution or incidence problem. Who benefits and who pays now occupy the attention of

---

* Resources for the Future, Inc. Special thanks are due to Allen V. Kneese and Elizabeth Duenckel of the RFF staff for advice, counsel, and much hard work. A note of appreciation is due to Eleanor B. Steinberg, who convinced me not to publish an earlier draft. Orris Herfindahl, Dennis Mueller, and Robert S. Steinberg read the manuscript and made helpful suggestions for improvement. I am also indebted to one of the reviewers for the Farquharson and Shapley references.

economists as they turn to questions of public goods. Welfare economics has unearthed no answer to the question of inter-personal comparisons of utility in the absence of a social welfare function. The generation of such functions has been delegated to the political process, since it is the formal social choice mechanism. Yet economists have found few political scientists interested in undertaking the task of explaining precisely how the generation occurs. Lacking the explanation, economists have tried their hand at it themselves, notably in the work of James Buchanan and Gordon Tullock (1965) and Anthony Downs.

Another economist, Kenneth Arrow (1951, 1963) had already explained how it is *not* done and that explanation has had more impact than the attempts to explain how it *is* done.[1] In brief, Arrow was interested, as were the Founding Fathers, in determining collective or social choices (choices of public policies or choices among candidates for office) on the basis of individual (voter) preferences. Arrow set up four seemingly reasonable conditions which a social choice mechanism should be expected to meet and found that, as a general proposition, no such mechanism could be devised. No one has disproved Arrow's Possibility Theorem, although some, particularly Duncan Black (Oct. 1969), have recently raised some hard questions about the relevance of Arrow's conditions.

Arrow had two parts to his Possibility Theorem and much more ink has been spilt about the General Possibility Theorem than about the Possibility Theorem for Two Alternatives. Arrow proved that the method of majority rule applied to two alternatives does satisfy the four conditions and that the Possibility Theorem for Two Alternatives was, "in a sense, the logical foundation of the Anglo-American two-party system" (1951, p. 48). The restricted Theorem has been largely ignored by later writers, perhaps because it was considered trivial from a social welfare function viewpoint.

Arrow did not pursue the implications of his restricted Theorem. His concern and the concern of most economists have been for the more general problem and, implicit in such concerns, the analogy with general equilibrium theory in economics. Such concerns and analogy are almost demanded by the logic of individual preference *orderings* and a social preference *ordering* by themselves.

Such orderings are incomplete postulates in political terms. For example, take the well-known voting paradox ordering

$$\begin{array}{ccc} A & B & C \\ B & C & A \\ C & A & B \end{array}$$

by which, with majority vote as the decision rule, no social ordering can be found. Implicit in that judgment is some voting matrix to which the preference orderings can be related. One such voting matrix is as follows (Y = yes vote, N = no vote, subscripts indicate ordinal ranking of ABC by each voter):

Issue	Voter		
	I	II	III
A	Y_1	N_3	Y_2
B	Y_2	Y_1	N_3
C	N_3	Y_2	Y_1

Here we assume each voter will vote for either his first or second choice, but never for his third choice. This assumption preserves the cyclical outcome for pairwise choices, A preferred to B, B preferred to C, and C preferred to A.

Suppose, however, another voting matrix which does no violence to the ordinal ranking of ABC by each voter:

---

[1] I am ignoring here the distinction made elegantly by Paul Samuelson between a Bergson social welfare function and Arrow's constitutional function.

Issue	Voter		
	I	II	III
A	Y_1	N_3	N_2
B	Y_2	Y_1	N_3
C	Y_3	N_2	Y_1

Now it seems clear that *B* will win, since Voter I is willing to vote for *A*, *B*, or *C*, and in so doing he provides the winning margin for either *B* or *C*. He prefers *B* to *C*.

Yet another voting matrix with the same ordinal[2] ranking:

Issue	Voter		
	I	II	III
A	N_1	Y_3	Y_2
B	Y_2	N_1	Y_3
C	Y_3	N_2	Y_1

Now it appears that Voters I and II will trade votes, defeating both *A* and *B* and allowing *C* to win. This results because both Voters I and II have, as a first choice, the *defeat* of an alternative. Thus, though *ABC* are mutually exclusive alternatives, trading is possible given the appropriate voting matrix.

These cases are commonplace occurrences in the political arena but they can not be examined by looking at preference orderings alone. In particular, the analogy of a market is particularly inapposite in the last example. There is no market mechanism by which I can directly express the intensity of my dislike for a product. The social choice mechanism of representative government combines a vote matrix with preference orderings to allow expression of negative intensities as well as positive intensities of preference. The use of the combined matrix will be shown later to be of value in understanding the utility mechanism.

Another reason for economists' growing interest[3] with collective or social choices is the increasing problem of externalities in the production process, particularly when these are harmful to the public. Air and water pollution resulting from man's economic activities are the most visible example. If air and water are free goods, or nearly so, then they may be overused because private cost calculations do not include the costs imposed on other people. Hence, as Buchanan and Tullock have noted,[4] the economic basis for taking collective actions has shifted from opportunities for external economies to conflicts over external diseconomies. The import of this shift is to focus attention on governmental action rather than on corporate action. For, while schemes can be devised, effluent charges for example, which would internalize these costs, they must be adopted to be effective. Adoption presumes a collective choice and persons hurt by this choice will resist it. If the scheme is rejected, how are we to judge the rejection? Is whatever the political process churns out "right" by definition? Some social scientists have called this trust of the political system into question. Daniel Bell asks for a new political theory to provide a way of choosing between the welfare gains and the welfare losses.[5]

---

[2] Note, however, that now the ranking must be strictly interpreted as intensity of preferences, whether for or against. Thus Voter I prefers first, *A* to lose; second, *B* to win; and third, *C* to win.

[3] See particularly the work of Robert Ayres and Allen Kneese and that of Lloyd Shapley and Martin Shubik.

[4] " . . . as people get richer, they need to rely less and less on their neighbors to cooperate in securing the individual benefits of possible joint activities, but they may need to rely more and more on some collective mechanism to prevent themselves, and their neighbors, from imposing mutually undesirable costs on each other. . . . 'congestion' replaces 'cooperation' as the underlying motive force behind collective action" (1965, p. 69).

[5] "The political tradition from John Locke to Adam Smith paved the way for a new society in which representative government and the free market economy served as the framework for a system of individual decision-making based on self-interest and rational choice. Can one write a new political theory . . . that deals with a service state and a society characterized by a new mixture of individual and communal public and private

Allen Schick warns that the political system may be defective in a manner analogous to an imperfect market. He notes that since the impact of public goods decisions falls unequally on different groups, the political mechanism, far from providing clear welfare criteria for choice, may produce either too much of a public good (defense?) or too little (environmental quality?). Is there no political analogue to Smith's invisible hand?

In sum, determining whether or not representative government does have a utility base and asking whether or not such a base is appropriate are both revelant areas for study. This paper is mainly concerned with demonstrating the utility base but does offer some thoughts on the latter question.

I. Representation and Individual Utility

The link between utility and representation has proved a major stumbling block which hinders our understanding of the utility base of representative government. That representative democracy, as opposed to pure democracy, was necessary for effective government was almost self-evident in the 18th century.

James Madison in Federalist Paper No. 55 expresses the point most succinctly, "Had every Athenian citizen been a Socrates, every Athenian assembly would still have been a mob." Need it be added that Madison's words relate to information costs, revealed preferences, and the lack of a vote-trading mechanism, or that modern proposals that everyone vote on all issues by electronic processes suffer the same defect as the Athenian assembly?

Arguing the necessity of representation on negative grounds does not address the question of the link between individual utility and the representative, however. Furthermore, the contention that the representative will be wiser, more judicious, less swayed by whims of the moment —whatever its truth may be—likewise begs the question. Let us address it directly.

Consider three men as comprising a district. Two independent issues are posited as being important for resolution, and the men's positions on these issues are described by the following combined vote and preference matrix:[6]

Issue	Voter		
	I	II	III
A	Y_2	Y_2	N_2
B	N_1	Y_1	Y_1

where again Y is a vote for, N a vote against, and the subscripts are the ordinal rankings of the issues by each man. In this case, all three men rank issue B as more important than issue A, Voter I prefers the defeat of issue B to passage of issue A, Voter II prefers the passage of B to the passage of A, and Voter III prefers the passage of B to the defeat of A.

Two methods exist whereby these men may decide these issues. They may meet as an assembly; in such case, it is obvious that under majority rule, both issues will pass (no trades are possible). Alternatively, they may elect a representative (not one of the three voters). In that case they face a mutually exclusive choice of one of four possible outcomes on the two issues. If we display these as alternative outcomes with the consequences for each voter under each outcome we have: (P = pass, F = fail)

---

decision-making units? . . . where joint decisions are to be made, are there clear welfare criteria that justify one choice rather than another?" (pp. 699, 977).

[6] In this and all subsequent examples, I assume no position on whether a representative should lead or follow his constituency. The individual preference orderings and voting positions can be considered sui generis or as having been formed from the persuasions of prospective or actual representatives.

Issue *A*	*P*	*P*	*F*	*F*
Issue *B*	*P*	*F*	*P*	*F*
Voter I wins	2	12	none	1
Voter II wins	12	2	1	none
Voter III wins	1	none	12	2

Thus, if a representative were elected on a $\left[\begin{smallmatrix}P\\P\end{smallmatrix}\right]$ platform, Voter I would win only his second choice (passage of *A*), Voter II would win both his first and second choices, and Voter III, his first choice only. Obviously, Voter II would vote for a representative espousing a $\left[\begin{smallmatrix}P\\P\end{smallmatrix}\right]$ platform, but Voters I and III would be motivated to look for alternatives to $\left[\begin{smallmatrix}P\\P\end{smallmatrix}\right]$. No single alternative is preferred by both voters. We may conclude that $\left[\begin{smallmatrix}P\\P\end{smallmatrix}\right]$ would win over any rival, and that the perceptive aspirant for office will run on this platform.

Sometimes an assembly outcome depends on vote-trading in which issues of lesser utility are traded for issues of greater utility. Consider this rearrangement of preferences:

	Voter		
	I	II	III
A	Y_2	Y_1	N_1
B	N_1	Y_2	Y_2

The assembly outcome of this arrangement would be that issue *A* passes and issue *B* fails. This results from an attempted trade between Voters I and III (which would result in failing both issues), which Voter II prevents by giving up issue *B* (voting *N* instead of *Y*), thereby keeping Voter I's *Y* vote on issue *A*.

Under an election process, obviously the voters cannot directly trade votes. They are again faced with a choice of mutually exclusive alternatives. The alternative outcomes are:

Issue *A*	*P*	*P*	*F*	*F*
Issue *B*	*P*	*F*	*P*	*F*
Voter I wins	2	12	none	1
Voter II wins	12	1	2	none
Voter III wins	2	none	12	1

Now, if we are to arrive at the same outcome $\left[\begin{smallmatrix}P\\F\end{smallmatrix}\right]$ as did the men when meeting as an assembly, the path which the choice process takes becomes crucial. Proceeding as we did before, Voters I and III can be attracted to $\left[\begin{smallmatrix}F\\F\end{smallmatrix}\right]$ as an alternative to $\left[\begin{smallmatrix}P\\P\end{smallmatrix}\right]$. Were $\left[\begin{smallmatrix}F\\F\end{smallmatrix}\right]$ to be chosen by one candidate, a second candidate could win on a $\left[\begin{smallmatrix}P\\F\end{smallmatrix}\right]$ platform, *assuming no other candidate runs.*

While it is not remarkable that a path can be found, by means of a chain of assumptions, which brings us to the solution reached through assembly trading, it is significant that the chosen path is not unlike a two-party system groping toward positions on issues.

Before formulating rules by which that path can be specified when more than two issues are involved, it may prove helpful to explore outcomes of all ordinal permutations of the $\left[\begin{smallmatrix}YYN\\NYY\end{smallmatrix}\right]$ voting matrix. These are shown in Table 1.

If we view each permutation as a separate district, we can see what happens when voters in different districts are not concerned about the *same* issues.

Suppose that the voters from districts (permutations) 4, 5, and 6 are now considered in terms of three issues, as follows:

	Issue	Voter		
		I	II	III
District 4	*A*	Y_1	Y_2	N_2
	B	N_2	Y_1	Y_1
	C	3	3	3
District 5	*A*	Y_2	Y_1	N_1
	B	3	3	3
	C	N_1	Y_2	Y_2
District 6	*A*	3	3	3
	B	Y_2	Y_2	N_1
	C	N_1	Y_1	Y_2

If an issue is not relevant in a district—as, for example, issue *C* in district 4—we can ascribe to it the third position in the ordinal ranking of each voter even though we ascribe no *Y* or *N* position. Since we

TABLE 1—LEGISLATIVE-ASSEMBLY OUTCOMES
2x3 Matrix

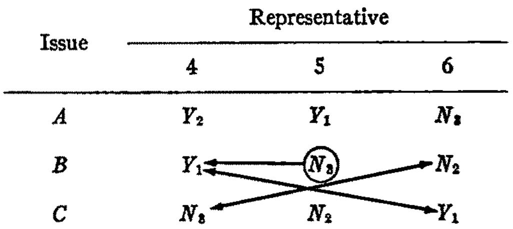
*Case 12	Voter			Outcomes	
Issue	I	II	III	Assembly	Representative
A	*Y*	*Y*	*N*		
B	*N*	*Y*	*Y*		
	Ordinal permutations:				
(1)	1	1	1	Pass	Pass
	2	2	2	Pass	Pass
(2)	1	2	1	Pass	Pass
	2	1	2	Pass	Pass
(3)	1	1	2	Pass	Pass
	2	2	1	Pass	Pass
(4)	1	2	2	Pass	Pass
	2	1	1	Pass	Pass
(5)	2	1	1	Pass	Pass
	1	2	2	Fail	Fail
(6)	2	2	1	Fail	Fail
	1	1	2	Pass	Pass
(7)	2	1	2	Pass	Pass
	1	2	1	Pass	Pass
(8)	2	2	2	Pass	Pass
	1	1	1	Pass	Pass

Note: *Case designations are formed by counting the frequency with which voters appear in the initial coalitions. In Case 12, the only nontrivial 2x3 case, one voter is in two coalitions, and two voters are in one coalition, hence Case 12.

have not changed the voting pattern or the ordinal rankings, we know the outcome in each district (whether decided by election of a representative or by assembly action) by reference back to Table 1. Thus the representatives from each district, when they meet in a regional assembly, could be shown as

Issue	Representative		
	4	5	6
A	Y_2	Y_1	3
B	Y_1	3	N_2
C	3	N_2	Y_1

If we use majority rule as the decision rule, then the blanks may be replaced, plausibly, by *N*'s, since the outcome is not changed thereby. The resulting matrix



Issue	Representative		
	4	5	6
A	Y_2	Y_1	N_3
B	Y_1	N_3	N_2
C	N_3	N_2	Y_1

has one potential trade (which is blocked by Representative 5) and the outcome is to pass the first two issues and fail the third.

Were the nine voters to come together as an assembly, thus

Issue	Voter								
	I	II	III	IV	V	VI	VII	VIII	IX
A	Y_1	Y_2	N_2	Y_2	Y_1	N_1	3	3	3
B	N_2	Y_1	Y_1	3	3	3	Y_2	Y_2	N_1
C	3	3	3	N_1	Y_2	Y_2	N_1	Y_1	Y_2

and the blanks replaced by *N*'s

Issue	Voter								
	I	II	III	IV	V	VI	VII	VIII	IX
A	Y_1	Y_2	N_2	Y_2	Y_1	N_1	N_3	N_3	N_3
B	N_2	Y_1	Y_1	N_3	N_3	N_3	Y_2	Y_2	N_1
C	N_3	N_3	N_3	N_1	Y_2	Y_2	N_1	Y_1	Y_2

again the first two issues pass and the third fails, although the process of trading which accomplishes this outcome is a little more complex. All issues are losing initially, but Voter VII must change his vote on issue *A* to block a trade between Voters I and III which would cause issue *C* to win. Similarly, Voter IV must change his vote on issue *B* to keep issue *C* from winning because of a possible trade between Voter III and either Voter V or VI.

While the foregoing illustration may be obvious, it does demonstrate that the process works across districts without uniformity of issues in every district.

Rules for Solution in the 3x3 Vote Matrix

Following William Riker in using only minimum winning coalitions, only two 3x3 vote patterns need be considered:

Case 030			Case 111		
Y	Y	N	Y	Y	N
Y	N	Y	Y	Y	N
N	Y	Y	Y	N	Y

Each case has 216 ordinal permutations. (Manipulation of Y's and N's are not necessary to exhaust the set of permutations.)

Legislative Vote Trading

Rules for solving the cases considered as assemblies are fairly straightforward. We assume voting stances are known but that relative importance of issues (ordinal ranking) is revealed only by actions. The rules are:

1) Trades take place if, and only if, they are mutually advantageous.

2) Any trader prefers a higher gain to loss ratio to a lower one, e.g., will trade a third choice for a first choice in preferences to a second choice for a first.

3) Any trade can be reversed (cancelled) by a third voter if he can offer one of the traders an alternative which is more advantageous to that trader and less damaging to himself than the trade would be.

4) Bluffs and threats, defined as actions which, if taken, would harm the actor, are not allowed.

5) While only ordinal utilities are used, it is convenient to set choice 1 = choices 2+3 for all voters.

6) Majority rule is the decision rule.

Solutions determined by these trading rules were assumed "correct" for the purpose of devising rules for selection of a representative.

Rules for Selecting a Representative

Again we assume that voting stances (opinion polls?) are known at the outset. The rules determine the decision path:

1) Start at the nominal outcome, i.e., count the votes. As presented here, that is always at the vector $\begin{bmatrix} P \\ P \\ P \end{bmatrix}$

2) Select from the remaining seven possible vectors, the outcome(s) which is (are) most advantageous:

(a) to the two voters "worst off" in the $\begin{bmatrix} P \\ P \\ P \end{bmatrix}$ vector[7]

(b) or, if the two "high" men are tied, to the one low man and either one of the high men[8]

---

[7] Example: Given a vote and ordinal matrix

Y_2	Y_1	N_1
Y_3	N_2	Y_2
N_1	Y_3	Y_3

hence possible outcomes

	P	P	F	F	P	P	F	F
	P	F	P	F	P	F	P	F
	P	P	P	P	F	F	F	F
Voter I wins	23	2	3	—	123	12	13	1
Voter II wins	13	123	3	23	1	12	—	2
Voter III wins	23	3	123	13	2	—	12	1

Voters I and III are "worst off." The only vector which is mutually advantageous is $\begin{bmatrix} F \\ P \\ F \end{bmatrix}$ (If more than one vector is possible, all are chosen.) Keep in mind that the ordinal ranking measures the importance of the issue to the voter. Thus $\begin{bmatrix} Y_2 \\ Y_3 \\ N_1 \end{bmatrix}$ is interpreted that the voter prefers first the defeat of C, second the passage of A, and third the passage of B. If the outcome is $\begin{bmatrix} P \\ P \\ P \end{bmatrix}$ this voter wins his second and third choices and not his first.

[8] Example:

P	P	F	F	P	P	F	F
P	F	P	F	P	F	P	F
P	P	P	P	F	F	F	F
13	3	1	—	123	23	12	2
13	123	3	23	1	12	—	2
23	3	123	13	2	—	12	1

The vector selected would be $\begin{bmatrix} F \\ P \\ F \end{bmatrix}$

(i) if more than one vector choice is possible, choose the one most favorable to the low man[9]

(ii) if no such vector exists, choose the outcome favored by the two high men[10]

(c) or, if all three voters are tied, to any pair of voters.[11]

3) If no vector can be chosen under (2), then the initial vector is chosen by all parties.

4) If a vector is selected under (2), then it becomes one party position.

5) One other party position is selected from the remaining six vectors by choosing the vector(s) which adds, and only adds, to the winnings of one member of the first party (in 2) above). If more than one outcome is possible, choose the outcome most advantageous to the voter not a member of the first party.[12]

6) The two party positions are matched and the vector selected which would receive the majority vote. (As set up here, this will always be the vector selected in 5) above.)

7) While only ordinal utilities are used, it is convenient to set choice 1 = choices 2+3 throughout for all voters.

8) Majority vote is the decision rule.

Results of applying these rules to the 3x3 matrix are shown in Table 2.[13] Several items in the results need some explanation. First, where a bargaining situation develops (either in the legislative process or in the selection process), the same assumption must be used if the two processes are to produce the same result. This is not surprising, but does mean that it is occasionally possible, particularly in Case 111, for some of the outcomes to be interpreted differently. Second, there are two permutations of Case 030 for which two solutions are possible because of our "choice 1 = choices 2+3" rule.

These two cases are the last vestige of the cyclical majority phenomenon. They are testimony to the sturdy truth underlying Arrow's General Therorem.

---

[9] Example:

P	*P*	*F*	*F*	*P*	*P*	*F*	*F*
P	*F*	*P*	*F*	*P*	*F*	*P*	*F*
P	*P*	*P*	*P*	*F*	*F*	*F*	*F*
23	2	3	—	123	12	13	1
13	123	3	23	1	12	—	2
13	3	123	23	1	—	12	2

The vector selected would be $\begin{bmatrix} P \\ F \\ F \end{bmatrix}$

[10] Example:

P	*P*	*F*	*F*	*P*	*P*	*F*	*F*
P	*F*	*P*	*F*	*P*	*F*	*P*	*F*
P	*P*	*P*	*P*	*F*	*F*	*F*	*F*
13	1	3	—	123	12	23	2
13	123	3	23	1	12	—	2
23	3	123	13	2	—	12	1

The vector selected would be $\begin{bmatrix} P \\ F \\ F \end{bmatrix}$

[11] Example:

P	*P*	*F*	*F*	*P*	*P*	*F*	*F*
P	*F*	*P*	*F*	*P*	*F*	*P*	*F*
P	*P*	*P*	*P*	*F*	*F*	*F*	*F*
23	2	3	—	123	12	13	1
23	123	3	13	2	12	—	1
23	3	123	13	2	—	12	1

The vector selected would be $\begin{bmatrix} P \\ F \\ F \end{bmatrix}$

[12] Example:

P	*P*	*F*	*F*	*P*	*P*	*F*	*F*
P	*F*	*P*	*F*	*P*	*F*	*P*	*F*
P	*P*	*P*	*P*	*F*	*F*	*F*	*F*
23	2	3	—	123	12	13	1
13	123	3	23	1	12	—	2
23	3	123	13	2	—	12	1

Vector $\begin{bmatrix} F \\ P \\ F \end{bmatrix}$ is chosen by one party. Two vectors dominate it under the add and only add rule: $\begin{bmatrix} P \\ P \\ F \end{bmatrix}$ and $\begin{bmatrix} F \\ P \\ P \end{bmatrix}$

The vector selected is $\begin{bmatrix} P \\ P \\ F \end{bmatrix}$

[13] All cases were solved by hand, but to avoid the possibility of a shifting premise, computer programs were written and the cases solved again. Fortunately, the two methods give identical answers. Elizabeth Duenckel developed the programs and her help is gratefully acknowledged.

TABLE 2—LEGISLATIVE-REPRESENTATIVE OUTCOMES
3x3 Matrix

*Case 030	Voter			Outcomes	
Issue	I	II	III	Legislative	Representative
A	*Y*	*Y*	*N*		
B	*Y*	*N*	*Y*		
C	*N*	*Y*	*Y*		
Ordinal permutations (216 possibilities):					
100 of which				Pass Pass Pass	Pass Pass Pass
38 of which				Pass Pass Fail	Pass Pass Fail
38 of which				Pass Fail Fail	Pass Fail Fail
38 of which				Fail Pass Pass	Fail Pass Pass
2 of which				Fail Fail Fail or Pass Pass Pass	Pass Pass Pass or Fail Fail Fail

*Case 111	Voter			Outcomes	
Issue	I	II	III	Legislative	Representative
A	*Y*	*Y*	*N*		
B	*Y*	*Y*	*N*		
C	*Y*	*N*	*Y*		
Ordinal permutations (216 possibilities):					
132 of which				Pass Pass Pass	Pass Pass Pass
42 of which				Pass Pass Fail	Pass Pass Fail
21 of which				Pass Fail Pass	Pass Fail Pass
21 of which				Fail Pass Pass	Fail Pass Pass

Note: *Cases are designated as in Table 1. Cases 030 and 111 are the only two nontrivial 3x3 cases.

II. The Cyclical Majority Uncycled[14]

While the paradox of the cyclical majority has been traditionally posed as involving mutually exclusive issues—alternative social states, candidates for one post, considered without side payments—both side payments and independent issues have been introduced into recent discussion. Arrow comments on side payments:

> If, instead of assuming that each individual votes according to his preferences, it is assumed that they bargain freely before voting (vote-selling), the paradox appears in another form. . . . If a majority could do what it wanted, then it would be optimal to win with a bare majority and take everything; but any such bargain can always be broken up by another proposed majority. [1969, p. 61]

We have seen, however, that once vote-trading is allowed, the paradox does not hold if issues are independent. That result is not limited to single-peaked preference functions, which Tullock and Paul Simpson have investigated; nor need it come about because of considerations of utility under uncertainty, as in the work of Coleman; nor does it depend on the existence of a pseudo-price mechanism, as posited by Robert Wilson.

To demonstrate this, let us examine the permutation usually considered in a mutually exclusive issue context as an independent issue case (using our rules for legislative trading). Changing the vote matrix to comply with the assumption that each voter votes for his first choice and against the other two, we will have (majority rule):

---

[14] James C. Coleman read an earlier version of this section and made valuable suggestions for improvement.

Issue	Voter			
	I	II	III	Outcome
A	Y_1	N_2	N_3	F
B	N_2	N_3	Y_1	F
C	N_3	Y_1	N_2	F

In order to examine the trading sequence, it will be useful to display the subscripts (rankings) separately and construct a trading matrix:

Issue	Voter		
	I	II	III
A	1	−2	−3
B	−2	−3	1
C	−3	1	−2

Each column vector shows a voter's trading desires. Applying our rules for trading, the voter *can* trade any cell with a negative sign but he will, obviously, only trade lower preferences for higher. Thus in this display each preference vector shows a desire and ability to trade either a second or third choice for a first choice. Choosing a first trade arbitrarily since none is dominant, the trading sequence is illustrated in Table 3.

Since there is no familiar notation to show the trading sequence, it may be helpful to describe the process shown in Table 3. Starting in Tableau I, and recalling that columns represent voters and rows are issues, we see that all three issues fail initially. Voters II and III trade votes on issues B and C (shown by the crossed arrows). This trade, if allowed to hold, results in the solution S_1, namely that, Voter I would win nothing, Voter II would win on his first and second preferences, and Voter III would win on his first and third preferences.

This trade is not stable, however, as it is in Voter I's best interest (and in his power) to reverse the trade by giving up his vote on issue C to Voter II so that Voter II does not need to trade with Voter III. Hence, the S_2 solution at the bottom of Tableau I.

Tableau II is generated by the S_2 solution, i.e., the issues are now F, F, P, and Voter I is shown voting Y on issue C. The negative and positive signs in the trading matrix are changed accordingly and additional trades are sought. There is a possible trade (again shown by crossed arrows) and this time the potential reversal cannot occur. Voter II could reverse this trade only if he is willing to give up Voter I's support on issue C. But, since issue C is his first choice, he will not be willing. Hence, S_3 is shown as the result of the trade between Voters I and II on issues A and B.

Tableau III is generated from this solution (S_3); all issues are now passing, and the new trading matrix shows no further trades are possible. Three votes are shown "circled," indicating that they are cast Y as the result of trades. The result does not depend on Voter I being willing to trade off *both* issues B and C to get A. In other words, $A \not> B+C$. He trades off C to keep B. When he has B, he can use it to get A, which he prefers to B.

When the solution $\begin{bmatrix} P \\ P \\ P \end{bmatrix}$ has been reached, it is stable unless the rule "choice $1 = $ choice $2+3$" is relaxed. If at least two voters think choices $2+3>1$, the solution could be forced to $\begin{bmatrix} F \\ F \\ F \end{bmatrix}$.

When this permutation appears in a mutually exclusive context (representative rules) the same two solutions appear. Displaying the outcomes as before:

P	*P*	*F*	*F*	*P*	*P*	*F*	*F*
P	*F*	*P*	*F*	*P*	*F*	*P*	*F*
P	*P*	*P*	*P*	*F*	*F*	*F*	*F*
1	12	—	2	13	123	3	23
1	13	12	123	—	3	2	23
1	—	13	3	12	2	123	23

and noting that $\begin{bmatrix}F\\F\\F\end{bmatrix}$ is the initial vector, we can choose no other vector under rule 2. Hence (under rule 3) $\begin{bmatrix}F\\F\\F\end{bmatrix}$ is chosen by both parties. Only if choice $1 > \text{choices } 2+3$ could $\begin{bmatrix}P\\P\\P\end{bmatrix}$ dominate $\begin{bmatrix}F\\F\\F\end{bmatrix}$

Thus the legislative rules choose $\begin{bmatrix}P\\P\\P\end{bmatrix}$ and the representative rules choose $\begin{bmatrix}F\\F\\F\end{bmatrix}$ This divergence is uniquely determined by the convenience rule, choice $1 = \text{choices } 2+3$. If this rule is relaxed, the two processes

TABLE 3—TRADING SEQUENCE

	Vote Matrix				Trading Matrix			Winning Preferences				
	Voter I	Voter II	Voter III	Out-come	Voter I	Voter II	Voter III			Voter I	Voter II	Voter III
						Tableau I						
Issue A	*Y*	*N*	*N*	*F*	1	−2	−3	Initial Solution	S_0	2nd 3rd	2nd 3rd	2nd 3rd
B	*N*	*N*	*Y*	*F*	−2	−3	1	(trade)	S_1	none	1st 2nd	1st 3rd
C	*N*	*Y*	*N*	*F*	−3 →	1	−2	(reverse)	S_2	2nd	1st 2nd 3rd	3rd
						Tableau II						
Issue A	*Y*	*N*	*N*	*F*	1	−2	−3					
B	*N*	*N*	*Y*	*F*	−2	−3 ↛[a]	1	(trade)	S_3	1st	1st	1st
C	Ⓨ	*Y*	*N*	*P*	3	−1	2					
						Tableau III						
Issue A	*Y*	*N*	Ⓨ	*P*	−1	2	3					
B	Ⓨ	*N*	*Y*	*P*	2	3	−1	Final	S_F	1st	1st	1st
C	Ⓨ[b]	*Y*	*N*	*P*	3	−1	2					

Note: ⤧ indicates trade. → indicates a reversal of trade. ↛ indicates a blocked reversal.

Each successive tableau of voting matrix and trading matrix is set up on the basis of trading done in preceding tableau; thus Tableau II starts at Solution S_2 and Tableau III at S_3.

[a] Reversal blocked because Voter I can retaliate by withdrawing his support on Issue **C** (last row).

[b] This vote remains a *Y* because it enables Voter I to hold Issue **B** (second row) so that he can trade it to Voter III.

chose the same outcome. If choice 1 $>$choices $2+3$, $\begin{bmatrix} P \\ P \\ P \end{bmatrix}$ is chosen. If choice 1$<$choices $2+3$, $\begin{bmatrix} F \\ F \\ F \end{bmatrix}$ is chosen. This convergence requires an addition to the representative rules to deal with the following special case when it occurs.

$$\left[\begin{array}{c} P \\ P \\ P \\ \hline 1 \\ 1 \\ 1 \end{array}\right]\left[\begin{array}{c} F \\ F \\ F \\ \hline 2\ 3 \\ 2\ 3 \\ 2\ 3 \end{array}\right]$$

A more general solution to this problem would require additional specified elements in each ordinal ranking. That was not attempted.

The other nominally cyclical examples (there are a total of 12 in each 216 permutations) come to a single stable solution both as independent issues (legislative rules) and as mutually exclusive choices (representative rules) and are considerably less complicated than the example presented here.

A caveat from the real world should be issued at this point. The practice of statecraft cannot always depend on legislative vote-trading or a two-party choice process. When decisions must be made rather separately, so that vote-trading possibilities are reduced, the decision is rarely made by simple majority vote. The role of the nominating committee as a screening device is familiar to all.[15] It was not by accident that nominating committees, party caucuses, the King's Council, and other devices grew up in Anglo-Saxon government to prevent indecision (lack of convergence) in governmental processes. This growth took place, not in the grip of utility theory as did the "democratic" elements of the system, but as a historical reaction developed during the long and bloody struggle for control of the English Crown. The early history of the English Commons shows it assenting to or rejecting a proposed levy. As the initiative gradually passed from King and Council to Commons, the utility part of the decision process was confined to adjustments on local issues. Individual members' preferences do not shape the bills involving national (rather than local) interests. The latter bills were (and are) shaped by party leadership.

W. Bagehot, writing near the middle of the last century, is still instructive on the point:

> . . . the principle of Parliament is obedience to leaders. . . . The penalty of not doing so is the penalty of impotence. It is not that you will not be able to do any good, but *that you will not be able to do anything at all.* If everybody does what he thinks right, there will be 657 amendments to every motion, and none of them will be carried or the motion either. [p. 141]

Moreover, there are times when even party leadership will not suffice. Writers from Aristotle—"the nature of a *polis* is to be a plurality"—to Coleman[16] have recognized that single issues considered *in vacuo* cannot be resolved by political means.

The knowledge that simple majority rule or individual utility concepts cannot be used in *all* matters of statecraft is no different, in principle, than the knowledge that the market cannot be trusted to establish a competitive price in a monopoly situation.

---

[15] Lewis Carroll's struggles with majority rule in the election of new officers at Christ Church College, Oxford, are recounted by Black (1963). C. P. Snow in *The Masters* gives an accurate picture of the undemocratic reality.

[16] ". . . there is evidence to suggest that when a single decision dominates a political or social system, . . . the decision process breaks down; and not only is there no "social welfare function," there is overt conflict . . . " (p. 1116).

III. Political Parties and Personal Utility

Earlier, Arrow's statement that the logical foundation of the Anglo-American two-party system could be found in his Theorem for Two Alternatives was mentioned in connection with the neglect with which the Restricted Theorem has been treated. It should now be clearer why the theorem is not trivial. If the two alternatives are not randomly chosen, but rather are those two positions, which, when put to a vote will result in the same choice as that chosen by the voters if they engage in vote-trading; *then a method of passing from individual tastes to social preferences, excluding inter-personal comparisons of utility, defined for a wide range of sets of individual orderings, neither dictatorial nor imposed, is representative government with a "satisfactory" two-party system.*

The word satisfactory in the above statement refers to whether or not the party system does tend to choose the two positions referred to above. Judging that is no simple task. Just as the price of gasoline at the corner station is no test of the efficacy of our market economy, neither is one issue, one time period, or one area a test of whether a two-party system is correctly reflecting utilities.

The proposition, and the rules on which it is based, have some strong implications. First, they suppose two parties and only two. For whatever reasons adopted, the single member, single vote constituencies in the United States exert a powerful force in that direction in any district or state, while the state electoral system of voting for president is a strong force for the two parties to be nationwide.[17] Second, they suppose non-doctrinaire parties, capable of changing positions to win voter approval. While often deplored (but not by politicians), the American party system qualifies on that count. Third, they suppose only a limited ability of parties to change course once committed in a particular election and less than total information of people's interests. These qualities are approximately present in nature. Fourth, they suppose majority rule to have utility for individuals. A recent discussion and proof of this proposition is given by Douglas Rae and Michael Taylor. Fifth, and finally, they suppose that the results from the permutations in the 3x3 cases are indicative of more general results. No proof is established of this supposition, although it should not be assumed that the set of issues which is important for voter preferences is much larger than three. Regression equations which explain voter and legislator behavior typically contain no more than four or five significant independent variables.[18]

The proposition and rules for electing a representative are in sharp contrast to any form of proportional representation, a system for reflecting every sizable shade of opinion in the legislature.[19] Yet the rules produce the same outcome as would occur if *everyone* were in the legislature.

IV. Optimality Considerations

If trading on independent issues and selection of candidates by their stand on issues are to proceed along optimal lines, then the issues must be "correctly" specified. Any issue, say, federal aid to education, may be framed in hundreds of different ways. How it is framed determines not only which people are for it and which against (specifying the vote matrix), but also the intensities of feeling pro and con (thus providing an input to the ordinal matrix). Control over how issues are to be framed is a powerful lever; one that is almost analogous to controlling the initial distribution of income in a market equilibrium analysis.

---

[17] Contrary to popular and Congressional opinion, the proposed change from awarding electoral votes on a winner-take-all basis, state-by-state to a simple nationwide vote count may greatly imperil the two-party system at the national level.

[18] See the work of Gerald Kramer and of John Jackson.

[19] See Duncan Black (1969) and the work of Ruth Silva for modern treatments of this perennial issue.

Let us illustrate this problem by first examining an ordinal matrix in which a legislature of five members is considering Issue A. Each member has his own version (bill) on this issue, as follows:

Issue	Members				
	1	2	3	4	5
A_1	1	2	3	5	4
A_2	2	1	2	4	5
A_3	3	3	1	3	3
A_4	5	4	5	1	2
A_5	4	5	4	2	1

It is not possible to solve this matrix without implicitly assuming some vote matrix. Before specifying the vote matrix, however, note the possibility of partitioning this matrix on affinity lines (which could correspond to party lines, liberal-conservative lines, urban-rural lines):

Issue	Members				
	1	2	3	4	5
A_1	1	2	3	5	4
A_2	2	1	2	4	5
A_3	3	3	1	3	3
A_4	5	4	5	1	2
A_5	4	5	4	2	1

Members 1, 2, and 3 show an affinity to each other's bills and an aversion to the bills of members 4 and 5. The affinity and aversion in this case are reciprocated. Now to solve this matrix, we must ask only whether the vote matrix of the upper left corner has one or more rows of Y's. If it has only one such row (say, A_2), then this version of the bill will dominate the matrix. Suppose, however, the vote matrix of this partition to be: (where the Y subscripts indicate ordinal ranking)

Issue	Members		
	1	2	3
A_1	Y_1	Y_2	Y_3
A_2	Y_2	Y_1	Y_2
A_3	Y_3	Y_3	Y_1

Pure bargaining appears to be indicated here, unless these three members belong to one party, and a party caucus under established rules of selection (majority vote, for example) is used to determine the outcome.

We cannot ignore members 4 and 5, however. If they are prepared to vote Y on their third choice, then the game is up and A_3 will dominate the matrix in the absence of pressures external to our consideration (party loyalty, for example). Is A_3 the "right" choice?

To examine that question, look at the total combined vote and ordinal matrix as specified by our assumptions on voting:

Issue	Members				
	1	2	3	4	5
A_1	Y_1	Y_2	Y_3	N_5	N_4
A_2	Y_2	Y_1	Y_2	N_4	N_5
A_3	Y_3	Y_3	Y_1	Y_3	Y_3
A_4	N_5	N_4	N_5	Y_1	Y_2
A_5	N_4	N_5	N_4	Y_2	Y_1

First, although A_3 is unique in having a unanimous Y vote, that fact is not significant. Suppose members 1 and 2 vote N on their third choice; A_3 still dominates the matrix. The crux of the matter is that only member 3 is in all minimum winning coalitions and he chooses A_3.

Still delaying an answer to the question, is A_3 the right choice, let us explore the general solution of these matrices with mutually exclusive alternatives through a series of logical statements:

1) Any ordinal matrix of n voters which displays different versions of one bill A

can be reduced to an nxn matrix if there are n different first choices.

2) If there are fewer than n first choices, the matrix can be reduced to a kxn matrix where $k<n$, where $k=1, \ldots n-1$.

3) Any such nxn and kxn ordinal matrix will have vote matrices which can be combined with it. Using majority vote as the decision rule, only rows containing at least $(n+1)/2$ (if n is odd) or $(n/2)+1$ (if n is even) Y votes need be considered.

(If no row has a majority of Y votes, there is no version of a bill on the issue A which can be passed by the legislature.)

(If only one row has a majority of Y votes, then this is the only version which can be passed.)

4) If two or more rows have (at least) a majority of Y votes, selection among them takes the following form:

- (a) for kxn matrices (every row passing),
  - (i) any minimum winning coalition (MWC) composed *only* of first choices is dominant (there can be no more than one such coalition in any ordinal matrix).
  - (ii) if no dominant row exists, then any rows with one or more first choices in an MWC should be compared. If there are common members of these coalitions, the common members will determine the solution. If the ordinal matrix of such common members, when reordered so as to put those members' highest preference on the main diagonal, results in a symmetrical matrix, the solution may be indeterminate.
- (b) for nxn matrices (every row passing),
  - (i) partition the ordinal matrix to include only members who are in two or more MWCs. If only one such member exists, his choice dominates.
  - (ii) if two or more such members exist, reorder the ordinal matrix to put their highest preference on the main diagonal. If the resulting ordinal matrix is symmetrical, the solution may be indeterminate.
  - (iii) if the resulting matrix is not symmetrical, the common members (by bargaining, caucus vote, or whatever) dominate the nxn matrix.

It is worth noting that ordinal symmetry in either the kxn or nxn matrix denotes a possible cyclical (indeterminate) case. Here, however, the meaning of the cycle is clear and its lack of decision benign. It denotes a lack of minimum agreement on an issue and hence chooses, correctly, to pass nothing. The issue is excluded from resolution pending a new set of legislators or a reformulation of the issue which can attract a better clustering of interests.

The same four logical statements can be used to describe the actions of a majority party in a legislature (if one assumed a high degree of party discipline) or of a committee system or dominant coalition of any kind. To take any number smaller than n, however, upsets the notion of majority rule; a notion which both Rae and Taylor have shown to have a strong claim on our rational interests. Thus, such decisions have some claim to be the "right" ones, and the presence of restrictions (committee dominance, for example) which exclude some members from the decision on how an issue is framed can be suspected of turning up with "wrong" decisions even though such exclusion is a way to mask indeterminate (cyclical) matrices.

Forming issues in an election, as opposed to formulating a bill in a legislature, is a far more imprecise process. A candidate is, properly, less concerned with each issue than in the design of a package of issues and a stance of each which will win over his opponent's package. The theoretical work of Otto Davis, Melvin Hinich, and Peter Ordeshook is particularly useful in defining candidate strategies and social welfare under various assumed distributions.

A related but separate question which has interested many writers, notably Black (1963) and Robin Farquharson, is the order in which votes are taken on bills which are interrelated. Unless one accepts party leadership as a guide both for candidate strategy and for legislative scheduling of voting, both areas can fall into indeterminancy under certain patterns of preferences.

V. Conclusions

Dennis Mueller concisely expressed the concern of many economists about the efficacy of vote-trading,

> . . . when voters are able to make and keep vote-trading agreements, their welfare will be greater than if no agreements were made. On the other hand, the resulting set of policy decisions will fall far short of being in any sense socially optimal. If the number of voters is not so large as to preclude the formation of partially stable coalitions, it is too small to remove completely the monopsony power a voter will be able to enjoy over any issue of vital importance to him. [p. 1310]

We have explored the role of party in the formation of proto-coalitions and vote-trading as the device for producing stability of outcomes while avoiding a coalition dominant on all issues. In so doing we reemphasize Madison's point (in Federalist Paper No. 10) regarding representative government as a defense against tyranny of the majority.

The monopsony problem is, however, a significant issue in representative government. Cases can be constructed in which, in a single legislature, one legislator with a strong interest in one bill can trade off many other votes to produce a majority for his bill. Notice, however, that to do so he must be in the number of minimum winning coalitions equal to the number of votes he needs on his bill. This is not an inconsiderable constraint, in theory or in practice.

An additional protection from monopsony power is the bicameral legislature—if the districts of the two houses are correctly drawn relative to one another. The point here is not "one man, one vote" since the whole of utility analysis is based on this principle,[20] rather the point is that district lines must be drawn so that representative patterns are significantly different in the two houses. What is advocated strongly by the lower house representative of district *A* may be safely resisted by the upper house senator whose constituency includes districts *A*, *B*, *C*, and *D*. Should a majority of these districts be of the same mind as district *A*, is not the senator then an advocate also? He is indeed, but if the number of senators is sufficiently restricted (which, in most senates it is not), this event happens only when a considerable portion of the state's electorate is of this mind. The case is not then one of monopsony power, but simply represents an intense minority preference which in utility terms may be accommodated through vote-trading. A further check occurs through executive veto power; the executive being the only representative of the whole electorate. These matters are discussed further in Buchanan and Tullock (1962), while Shapley has explored the theoretical problem in his formulation of compound-simple games.

There are many peculiar "legislatures," special districts, commissions, and so forth, which exist as single houses without the check of a second house or of executive veto. The political scientist's intuitive mistrust of the use of these devices for public decision making is (or should be) rooted in the utility defect which these bodies possess. The defect occurs not only because they are prone to single dominant majorities but also because the two-party system does not operate in them. The common practice of having the Governor, or his appointee, sit on interstate agencies, for example, may be "political" but it has no connection to the utility concerns under discussion here. This practice grew up because it was administratively convenient and the issues, in the past, not so socially significant. To continue the practice now in agencies like interstate water resources commissions, or port authorities, when the decisions made by such agencies are social choices of the most critical nature, is not to be countenanced on any utility principle. Committee rule. seniority, and other twentieth century habits of general legislatures may be condemned on the same grounds. In the best of legislatures, however, it is difficult to conceive of a perfectly competitive vote market, with marginal utilities proportional to "prices." The use of the term price has no meaning. Perhaps all that can be said is that trades take place at the margin for each person, i.e., that any legislator will prefer paying a lower price (changing his vote on an item of lesser interest to him) rather than a higher price but will always be willing to trade so long as the ratio of gain to loss is above unity. While I have elsewhere demonstrated that the probability of trading increases as the number of independent issues increases, it does not follow that prices therefore tend to approach marginal conditions, for there remains no comparability among utilities.

---

[20] Note the work of J. R. Pole. It should be remembered that senates were traditionally designed to counter the individual utilities reflected in the lower houses.

REFERENCES

K. J. Arrow, *Social Choice and Individual Values*, 1951, 2d. ed., New York 1963.

———, "The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-Market Allocation," in *The Analysis and Evaluation of Public Expenditures: The PPB System*, Subcommittee on Economy in Government of the Joint Economic Committee of U.S. Congress, Washington 1969, pp. 47–63.

R. U. Ayres and A. V. Kneese, "Production, Consumption and Externalities," *Amer. Econ. Rev.*, June 1969, *59*, 284–97.

W. Bagehot, *The English Constitution*, 2d ed., London 1905.

D. S. Bell, "A Summary of the Chairman," *Daedalus*, summer 1967, *96*, 698–704, 975–77.

D. Black, *The Theory of Committees and Elections*, Cambridge 1963.

———, "Lewis Carroll and the Theory of Games," *Amer. Econ. Rev. Proc.*, May 1969, *59*, 206–10.

———, "On Arrow's Impossibility Theorem," *J. Law Econ.*, Oct. 1969, 227–47.

J. M. Buchanan and G. Tullock, *The Calculus of Consent: Logical Foundations of Constitutional Democracy*, Ann Arbor 1962.

——— and ———, "Public and Private Interaction under Reciprocal Externality," in J. Margolis, ed., *The Public Economy of Urban Communities*, Baltimore 1965.

J. S. Coleman, "The Possibility of a Social Welfare Function," *Amer. Econ. Rev.*, Dec. 1966, *56*, 1105–22.

O. Davis, M. Hinich, and P. Ordeshook, "An Expository Development of a Mathematical Model of the Electoral Process, *Amer. Polit. Sci. Rev.*, June 1970, *64*, 426–48.

A. de Tocqueville, *Democracy in America*, New York 1947.

A.. Downs, *An Economic Theory of Democracy*, New York 1957.

R. Farquharson, *Theory of Voting*, New Haven 1969.

L. K. Frank, "The Need for a New Political Theory," *Daedalus*, summer 1967, *96*, 809–16.

E. T. Haefele, "Coalitions, Minority Representation, and Vote-Trading Probabilities," *Publ. Choice*, spring 1970, *8*, 75–90.

M. Hinich and P. Ordeshook, "Plurality Maximization vs. Vote Maximization: A Spatial Analysis with variable Participation," *Amer. Polit. Sci. Rev.*, Sept. 1970, *64*, 772–96.

J. E. Jackson, "A Statistical Model of United States Senators' Voting Behavior," doctoral dissertation, Harvard Univ. 1968.

G. Kramer, "Short-Term Fluctuations in U.S. Voting Behavior: An Econometric Model," mimeo 1967.

D. C. Mueller, "The Possibility of a Social Welfare Function: A Comment," *Amer. Econ. Rev.*, Dec. 1967, *57*, 1304–11.

J. R. Pole, *Political Representation in England and the Origins of the American Republic*, New York 1966.

D. W. Rae, "Decision Rules and Individual Values in Constitutional Choice," *Amer. Polit. Sci. Rev.*, Mar. 1969, *58*, 40–56.

W. H. Riker, *The Theory of Political Coalitions*, New Haven 1962.

P. A. Samuelson, "Arrow's Mathematical Politics," in S. Hook, ed., *Human Values and Economic Policy*, New York 1967, 41–51.

A. Schick, "Systems Politics and Systems Budgeting," *Publ. Admin. Rev.*, Mar./Apr. 1969, *29*, 137–51.

L. S. Shapley, *Compound Simple Games*, Santa Monica 1967.

——— and M. Shubik, "On the Core of an Economic System with Externalities," *Amer. Econ. Rev.*, Sept. 1969, *59*, 678–84.

R. C. Silva, "Relation of Representation and the Party System to the Number of Seats Apportioned to a Legislative District," *Western Polit. Quart.*, 1964, *17*, 742–69.

P. B. Simpson, "On Defining Areas of Voter Choice: Professor Tullock on Stable Voting," *Quart. J. Econ.*, Aug. 1969, *83*, 478–90.

M. Taylor, "Proof of a Theorem on Majority Rule," *Behav. Sci.*, May 1969, *14*, 228–31.

G. Tullock, "The General Irrelevance of the General Impossibility Theorem," *Quart. J. Econ.*, May 1967, *81*, 256–70.

R. Wilson, "An Axiomatic Model of Log-Rolling," *Amer. Econ. Rev.*, June 1969, *59*, 331–41.

International Trade and Capital Mobility

By Ernest Nadel*

It has long been recognized that commodity movements and factor movements are, to a degree, substitutes for each other in international exchange (see Carl Iverson, Mountifort Longfield, James Meade, Bertil Ohlin, John H. Williams). Yet, until recently, the dominant theory of international trade, the Heckscher-Ohlin (H-O) model, had been rather thoroughly analyzed under the rigid assumption of the immobility of factors. Only in 1957, with the publication of Robert Mundell's important article, was capital mobility in a H-O model explored. This paper presents a fuller treatment of capital mobility in the H-O model. We discuss the model under conditions of tariffs on goods flows and taxes on capital relocations.

Nations may trade by exchanging goods or by exchanging their relatively abundant factors which produce those goods. We demonstrate that the substitutability between these two avenues of exchange continues to hold even under conditions of tariffs and taxes. The dynamics and equilibria are demonstrated, showing that a nation will pay for its imports *either* through exports of goods *or* through earnings on foreign-placed capital, *not* through *both* of these methods. Tariffs and tax rates will dictate which will occur, i.e., tariffs and taxes will be shown to affect the *pattern of trade*, not merely the quantities of commodities traded.

We also correct a hitherto general and unrecognized error. We show that the levying of a tariff does *not* necessarily generate a relative price differential (of final goods prices) between the trading countries equal to the tariff proportion. The relative price differential will often be less than the tariff proportion, and this holds even though the good is still imported into the country. Further, contrary to previous results (see Ronald Jones 1967), we show that, generally, a tariff-cum-tax levy will not result in the complete specialization of production.

I. Introduction to the Analysis—A Substitutability Theorem

In the standard H-O pure barter theory of international trade, the goods exports of one country are exchanged for the goods exports of the other. In his article, Mundell discussed the effect of capital mobility in such a model. Capital mobility is defined as a relocation of capital (machines) from one country to the other accompanied by a repatriation of foreign earnings. His analysis opened up a new area of analysis of exchange between nations. In the normal H-O model, a country exports the goods in whose production its relatively abundant factor is relatively intensively used. Mundell showed that the exact same free trade results, in terms of equilibrium price ratio, factor rewards, and consumption levels, can be achieved by a nation exporting its relatively abundant factor itself instead of exporting its final consumer good.

In this paper, we continue to consider the human factor of production, labor, to be immobile. (The immobility of all labor implies also the immobility of the locus of consumption in that owners of capital are themselves part of the labor force. Thus, the repatriation of foreign earnings can be viewed as an implication of the assumption of labor immobility, and not as a separate, additional assumption itself.)

---

* University of California, Berkeley.

We will employ the two-country (H, F), two-good (steel, cotton), two-factor (capital, labor) model, with constant returns to scale in production, exactly identical production functions in each country, and no reversals of factor intensity occurring. Steel is always capital-intensive, and country H is initially endowed with a relative abundance of capital.

The consideration of capital mobility in effect turns the model into a three-good model. In the analysis which follows, these three goods are: steel (the home country's "natural" export good under conditions of capital immobility); cotton (the home country's natural import good under conditions of capital immobility); and capital services (a once-and-for-all relocation of a stock of capital being equivalent to a continuous flow of the services of that capital).

Assuming repatriation of foreign earnings, we will discuss the three basic patterns of exchange between the two nations that may occur:

Pattern 1: Country H can *import cotton*, and pay for it by *exporting steel*. No capital services need be traded. (This could be shown by production at P in Figure 1; consumption is at point C.)

Pattern 2: Country H can *import cotton*, and pay for it by *exporting capital services*. No steel need be traded. (This could be shown by production at J in Figure 1, consumption at C.)

Pattern 3: Country H can *import steel*, and pay for it by *exporting capital services*. No cotton need be traded. (This could be shown by production at N in Figure 1, consumption at C.)

Combinations of these patterns are possible, such as importing cotton and paying for it by exports of steel and capital services. Our analysis will deal mainly with the simple patterns discussed.

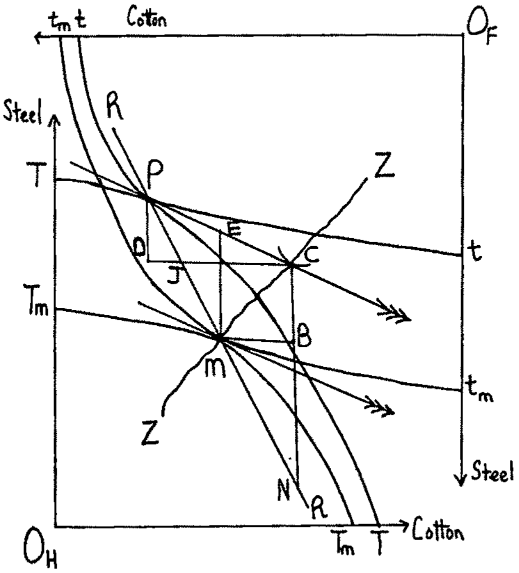


FIGURE 1—REPRESENTATION OF EQUILIBRIUM DURING FREE TRADE

Although our model is a three-good model, it will be shown that, except for razor's edge cases, only two of the three goods will be traded internationally. And, in those razor's edge cases, (i.e., when the tariff and tax levels happen to satisfy a certain relationship), all three goods could enter international trade, but it would not be necessary that they do so. Trade in only two goods will always suffice.

When analyzing the effects of tariffs and taxes, it will be shown that it is incorrect to *assume* that the post-interference trade *pattern* will be similar to the pre-interference pattern.

We will first briefly summarize the Mundell analysis demonstrating the substitutability between goods flow exchange and factor relocation exchange between nations. Then, we will use this "substitutability theorem" (our name for the central proposition, but the analysis essentially duplicates Mundell's), to investigate fully the consequences of interferences due to tariffs and taxes on foreign earnings.

With no domestic capital invested in the foreign country, free goods trade equilibrium will be as shown in the following diagram.

TT is country H's transformation curve. Production is at P, consumption at C. Country F's transformation curve is tt. Again, country F's production and consumption are also at P and C, respectively, considered relative to the origin, O_F, of country F. PD of steel from country H is exchanged for DC of cotton from country F.

RR is the Rybczynski contraction (expansion) line, for the price ratio given by the slope of PC. At that price ratio, if there is a capital outflow (inflow), the transformation curve of H contracts (that of F expands); in so doing, at its point of intersection with RR, the slope of the new transformation curve is equal to the slope of PC. Of course, as capital moves, say, from country H to country F, the two changing transformation curves continue to be tangent to each other along RR, with a slope that of PC. Production in both countries therefore will change in a southeasterly direction. World production is unchanged.

For barter trade with factor immobility, country H would export PD of steel in return for imports DC of cotton. If instead exchange were to take place via a relocation of capital, a solution would, as Mundell showed, be as follows. Let ZZ represent the Engel curve, for price ratio given by slope PC, for those consumers in country H who do send their capital to country F. The curve ZZ crosses the expansion line RR at M which is the final production point in the two countries. The boundary TT contracts to T_MT_M; similarily tt expands to t_mt_m. Let K_M represent the implicit volume of capital relocation. Therefore MB of cotton plus BC of steel will represent the interest earnings on K_M. Furthermore, MB of cotton plus BC of steel represents the exact composition of consumption these capital exporters desire because by construction of ZZ, the difference between C and M represents desired consumption of capital exporters, at unchanged goods and factor prices.

Thus, factor relocation can generate the exact same results as free trade in goods with factor immobility; the only differences are that the distribution of the location of capital among countries is different and that there is a continuous (one-way) repatriation of foreign earnings, instead of a continuous (two-way) goods flow. But, in terms of factor rewards (i.e., rentals on capital, which are not the same as rates of return on capital), relative prices, and hence consumption loci, the two methods of exchange are identical. Notice, too, that *in the factor relocation case no barter exchange occurs between the two nations, in the sense that imports are paid for out of foreign earnings, and not by an equivalent amount of final goods exports.* Aside from the repatriation of foreign earnings, the two nations are self-sufficient.

(The reason why the two nations can become self-sufficient is to be found in the Rybczynski theorem. Under free trade, country H produces more steel and less cotton than it desires; and vice versa for country F. The relocation of capital causes cotton production to increase and steel production to decrease in country H and vice versa in country F. These output changes adjust so as to satisfy domestic demands so that eventually the need for imports fall to zero.)

Effects of Tariffs and Taxes

We can now proceed to analyze the effects of tariffs and taxes. As already shown, goods flow and factor relocation are substitutes. It turns out that tariff and tax interferences create incentives which will dictate, via profit maximizing behavior, whether goods flow or factor relocation will take place. These incentives will determine which of the three patterns of exchange discussed above will actually occur. (To avoid irrelevant complications, we will assume that the government redistributes the tariff or tax proceeds back to its citizens.)

Beginning with a situation of free goods trade, we consider the effects of country H imposing a tariff (100 τ percent) on cotton imports. No tax is imposed on foreign earnings, and it is also assumed, *temporarily*, that purchases of foreign goods paid for out of repatriated earnings are *exempt* from the tariff applicable to commercial trade. As Mundell has shown, the tariff interference would raise the price of cotton in country H, so long as cotton is imported, i.e., so long as barter trade exists. And, by the Stolper-Samuelson theorem (see Wolfgang Stolper and Paul Samuelson), a higher price of cotton means a lower rental to capital. (The opposite price and factor reward movements occur in country F. The rental on capital is a unique, monotonic function of the relative goods price; because production functions are the same in both countries, this rental price function is the same in both countries.) Thus, there is an incentive for capital to move from country H to country F, i.e., capital relocation becomes preferable to goods trade. The capital outflow will cease when prices and hence rentals on capital are again equalized in the two countries. But, so long as barter trade exists, prices in the two countries will differ by the amount of the tariff. So, for rentals on capital to be equalized, barter trade must disappear; the two nations must become self-sufficient; we must reach the factor relocation solution.

One such solution (as shown by Mundell) is represented by production point M in Figure 1, which implicitly reflects the transfer of an amount of capital K_M.

Actually, if purchases of foreign goods paid for out of repatriated earnings are not subject to the tariff, any point along RR southeast of (and including) point J in Figure 1 is a possible equilibrium production point. This can be shown as follows. At unchanged terms of trade, with production at P, country H is willing to export some steel (PD) in return for some imports of cotton (DC). As capital moves from country H to country F, country H is able to purchase cotton out of its repatriated foreign earnings. Also, as the capital relocation occurs, production of steel falls and that of cotton rises in country H. Thus, excess demand for cotton falls. Eventually, as capital relocates, all the excess demand for cotton is exactly satisfied by the repatriated earnings. This occurs at point J. For a relocation of a quantity of capital equal to or greater than that implicitly represented by production point J (i.e., K_J of capital), there is no longer any need for country H to export steel to pay for desired cotton imports: i.e., there will no longer be any strict barter trade.

We now examine what occurs when the tariff on imported cotton goods also applies to the purchases of cotton goods out of repatriated earnings, i.e., when there is *no exemption*. This of course is what normally occurs when an import tariff on cotton is levied.

In this case, so long as cotton imports are coming into country H from country F, the tariff will be applied to them. If the tariff is effective, it creates a relative price differential between the two countries, hence a rental on capital differential as well. So long as this occurs, capital will continue to move from country H to country F. Capital will relocate between the two countries until there no longer are any imports of *cotton* into country H, since the tariff applies to imports of cotton only. This will occur when production point N is reached, reflecting a relocation of K_N of capital from country H to country F ($K_N > K_M$). Until production point N is reached (at unchanged terms of trade), there will always be a desire to import cotton into country H. But, by the Rybczynski Theorem, when point N is reached, the excess demand for cotton in country H falls to zero. At that point, NC of steel are the only imports into country H, NC of steel also being equal to the rental on K_N of capital placed in country F. When point N is reached, the old free trade equilibrium results are restored.[1]

Thus, even when τ applies to all cotton imports, a capital relocation can occur which will create a successful "evasion" of the tariff. The escape valve is a change of trading pattern, with country H becoming an importer of steel, rather than a exporter of steel. There is a switch from the first pattern of trade described in Section I to the third pattern. Thus, Mundell was correct in his paper when he suggested that so long as purchases out of foreign earnings are subject to the tariff, the free trade results will be disturbed. However, he did not consider the possibility of the "switching" of trade patterns described above. Of course, with the switching possible, purchases of steel out of foreign earnings are not subject to the tariff, because there is no tariff on steel.

This evasive switch of trading pattern is rendered impossible if the tariff applies to *all* foreign goods entering the home country—*cotton or steel*; it is also impossible if a tax is levied *directly upon foreign earnings*. We now turn to analyze these cases.

II. Conditions for Capital Relocation and Correction of a Fundamental Error

We must first consider the general relationship that obtains in this model between differences in relative prices between the two countries and differences in real rentals on capital between the two countries. This is the familiar factor price/goods price relationship implicit in the technology we are assuming, and a reflection of the Stolper-Samuelson theorem. This relationship holds so long as *specialization is incomplete* in the two countries.

Because we are assuming identical technologies in the two countries, a difference in relative prices in the two countries creates (or, is accompanied by) a difference in real rentals on capital in the two countries.

We shall adopt the following symbols:

P_C = nominal price of cotton (For simplicity, we may assume that both countries use the same nominal currency standard, e.g., dollars. The nominal price would then be the dollar price.)
P_S = nominal price of steel
$p = P_s/P_C$ = relative price of steel in terms of cotton
r = real rental on capital in terms of cotton
R = nominal rental on capital $= r \cdot Pc$
dx/x = the proportional change in the variable x

Making use of the magnification effect discussed by Jones (1965), $(dr/r)/(dp/p) = \alpha > 1$. (This technological relationship is expressed in terms of the good cotton. There is no loss in generality in doing so.) Furthermore, $dR/R = dr/r + dP_C/P_C$.

We begin with a situation of free trade and no home capital placed abroad. Country H initially exports steel in return for imports of cotton.

We will consider the case where the home country levies a tariff of 100τ percent on *all* imports, cotton and steel, into country H and also levies a tax of $100t$ percent on the earnings of any capital that is placed in the foreign country.

Now, the tariff will initially raise the price of cotton in country H.

---

[1] It is possible that position N could not be achieved, in that it may lie off the diagram, below the horizontal axis. In this case, specialization in cotton production would occur. Position M can always be reached.

$$\frac{dp}{p}=\frac{dP_S}{P_S}-\frac{dP_C}{P_C}=-\frac{dP_C}{P_C}=-\tau$$

(Initially, the price of steel does not change, i.e., $dP_S/P_S=0$.)

This in turn leads to $dr/r=\alpha dp/p=-\alpha\tau$, and $dR/R=dr/r+dP_C/P_C=-\alpha\tau+\tau=-\tau(\alpha-1)<0$, because $\alpha>1, \tau>0$.

For ease of exposition, we hold foreign prices and foreign rental on capital constant. (This is hardly crucial at this point, for what we are actually concerned with are *differences* in prices and rentals between the two countries.)

Because of the tax (100t percent) on foreign earnings, the rental that home capitalists can earn by sending their capital abroad changes by $-100t$ percent. But, the tariff causes $dR/R=-\tau(\alpha-1)$ in country H. The tariff-cum-tax combination will cause a capital outflow from country H to country F only if the fall in the nominal return to capital at home is larger than the decrease in the net return earned by home capitalists who would invest abroad;

$$\text{i.e., if}\quad -\tau(\alpha-1)<-t,$$

$$\frac{t}{\tau}<\alpha-1>0$$

Thus, we have derived the condition for an induced capital outflow consequent to levying a tariff-cum-tax combination. Capital will relocate from country H to country F if $t/\tau<\alpha-1$. (If $t/\tau=\alpha-1$, home capitalists will be indifferent as to where they place their capital. Hence here, and throughout the paper, we have used only the inequality conditions. In our example of the previous section, with $t=0$, we had $t/\tau=0/\tau<\alpha-1$, so that the tariff (-cum-tax) combination did indeed cause a capital outflow.)

We can also specify the equilibrium conditions that must hold when the capital relocation ceases. Net nominal rentals to home capitalists must be equal in the two countries. The gross foreign rental is held constant at its free trade level, by assumption. Thus, compared to free trade magnitudes, the rental to home capitalist investors abroad will change by $dR/R=t$. This must be equal to the change in the nominal rental to capital in country H.

Assume for now that only cotton is imported into country H. (Below, we will present the necessary conditions for this to occur.) Thus, $dP_C/P_C=\tau$, (i.e., the relative change in the price of cotton in country H, relative to its free trade level, and hence also relative to the unchanged price of cotton in country F, will be exactly equal to the tariff proportion.) Let dp^*/p represent the proportional change in the home relative price of steel that will have occurred once equilibrium is reached. (Thus dp^*/p also represents the final equilibrium relative price differential between the two countries.)

We again express the change in the home nominal rental:

$$\frac{dR}{R}=\frac{dr}{r}+\frac{dP_C}{P_C}=\frac{dp^*}{p}\cdot\alpha+\tau$$

In equilibrium, we must have

$$\frac{dp^*}{p}\cdot\alpha+\tau=-t,$$

i.e.,

$$\frac{dp^*}{p}=-\frac{t+\tau}{\alpha}$$

Summarizing, we have the following:

For a capital relocation to occur, we must have $t/\tau<\alpha-1$.

Once equilibrium has been restored (when the capital relocation has ceased), we must have $dp/p^*=-(t+\tau)/\alpha$.

Normally, it is assumed that a tariff of 100τ percent will cause a relative price differential in equilibrium of $dp^*/p=-\tau$. By our analysis, this implies $dp^*/p=-\tau$

$= -(t+\tau)/\alpha$. This implies $t/\tau = \alpha - 1$. However, this contradicts our conclusion that for a capital relocation to even occur we must have $t/\tau < \alpha - 1$. Clearly, an impasse has been reached.

We have come to the crux of a very important issue. If capital relocation is to occur, *and* incomplete specialization is also to be retained, *and* an equilibrium state is to be finally reached, the relative price differential will be equal to the tariff only if $t/\tau = \alpha - 1$. However, we have seen that the condition for a capital relocation to occur in the first place is $t/\tau < \alpha - 1$.

The problem can be demonstrated in another way. By the Stolper-Samuelson theorem, and given our assumptions about the production functions in the two countries, there is a unique relationship between relative price differentials and rental on capital differentials between the two countries, so long as specialization is *incomplete* in the two countries. When a country levies a tariff (100τ percent) and a tax ($100t$ percent), it seeks to generate a relative price differential of 100τ percent and a relative nominal rental on capital differential of $100t$ percent. It is only by chance that the tariff and tax levied will be related to each other in the *exact* proportion that the technology (i.e., the Stolper-Samuelson theorem) *necessitates*. If that fortunate situation does not occur, something in the system must "break down."

It was Jones (1967) who first recognized the existence of this impasse. Until his analysis, no writers[2] had worried about the constraints that technology (i.e., the Stolper-Samuelson theorem) imposes on the system when tariffs and taxes are levied. Jones' resolution of the difficulty was to state that it must be incomplete specialization which breaks down, one or both countries being forced to specialize in production. When this happens, the Stolper-Samuelson theorem no longer imposes the constraint of the magnification effect. With incomplete specialization gone, we can no longer require that in equilibrium $dp^*/p = -(t+\tau)/\alpha$.

Jones' resolution, although it does overcome the impasse, is not in general the correct one—i.e., market forces will not in general bring us to his solution. Instead, the resolution which we present not only resolves the impasse, it is also the solution which market incentives will dictate. (The Jones solution is correct, and is brought about by market forces only if steel is eventually imported into the country *and* the point N (Figure 1) lies off the diagram. However, with a uniform tariff on *all* imports, steel is never imported into the country. The Jones analysis, however, did *not* consider a *uniform* tariff levied on cotton *and steel* imports.)

We attempt to show that as a result of an induced capital relocation incomplete specialization does not break down. Instead, we show that relative goods prices in the two countries will not differ by the amount of the tariff—they will differ by *less*. This is rendered feasible because the pattern of international exchange is drastically altered. Before the capital relocation both goods, steel and cotton, enter into trade. In such an instance, relative prices must differ by the amount of the import tariff. But, because of the capital relocation, one of the final goods, steel or cotton as the case may be,[3] *no longer enters into international trade* (i.e., we change from pattern 1 of trade, to patterns 2 or 3, cf. Section I).

Of course, the *nominal* prices of the good that remains internationally traded do differ by the tariff magnitude as between countries. But relative prices do not have to differ by the magnitude of the tariff.

---

[2] One can safely say that practically all writers in trade theory have implicitly assumed the joint effectiveness of any combination of tariff and tax (except in the cases of totally prohibitive tariffs or tariffs on nonimported items).

[3] This depends on the relationship between the tariffs on steel and on cotton. With a uniform tariff, steel disappears from trade. See more detailed results below.

Essentially, the error others have made is to *assume* what the final pattern of international exchange will be like, i.e., similar to the pattern observed before capital mobility and the presence of tariffs and taxes are considered. What we show is that such an assumption is, in general, not valid. The pattern of trade must come out of the analysis—it must not be assumed. What we will show below is that the tariff and tax parameters will actually determine what that pattern will be like.

We will substantiate these propositions more fully below. At this point we will present a diagrammatic summary of our results.

In Figure 2 we have drawn a line with slope $\alpha-1$. The locus of positive t and τ that country H can set is the entire (first) quadrant. For any combination of t and τ lying above the line with slope $\alpha-1$, capital relocation will not occur. For any combination of t and τ lying below the line, capital relocation will occur (because $t/\tau<\alpha-1$).

Let V be one such latter point, representing a combination of a tariff of $100\tau_v$ percent and a tax of $100t_v$ percent which will lead to capital relocation. We can show what the final equilibrium relative price dp_v/p will be when the capital relocation has stopped, i.e., $dp_v/p=(t_v+\tau_v)/\alpha$.

To show dp_v/p on the figure, draw a line through V having a slope of $-45°$. That line crosses the line $t/\tau=\alpha-1$ at A. The abscissa of point A is $dp_v/p=(t_v+\tau_v)\alpha$ $(=CB=OE)$.

Also notice that $dp_v/p=OE<OD=\tau_v$, i.e., $dp_v/p<\tau_v$.

We can also show another interesting result. We can show that the change in the *real* return to capital is *greater* than $100t$ percent. If we measure the change in the real return to capital in terms of cotton,

$$\left(\frac{dr_c}{r_c}\right)=\alpha\left(\frac{dp}{p}\right)_v=EF\text{ (on the figure) }>t_v$$

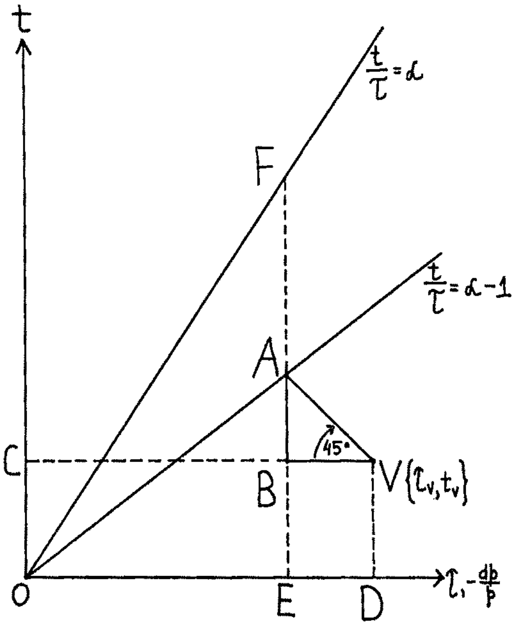


FIGURE 2—EQUILIBRIUM MAGNITUDES AFTER CAPITAL RELOCATION OCCURS

Thus, if capitalists consume only cotton, their real return falls by an amount greater than the tax on foreign earnings. And, even if they consume only steel, we get

$$\frac{dr_s}{r_s}=(\alpha-1)\left(\frac{dp}{p}\right)_v=EA>t_v$$

Thus, no matter what their consumption pattern is like, the real return to capital will fall as a result of the tax-cum-tariff combination, and it will fall by *more* than the rate of tax on foreign earnings.

III. Equilibrium After Factor Relocation

Algebraic Presentation

Before we derive any further results we must demonstrate that what we said happens actually does happen, i.e., we must show that, when $t/\tau<\alpha-1$, incomplete specialization does not disappear, but instead the induced capital relocation causes a change in the trading pattern, country H no longer exporting steel, and the relative price differential between the two countries being less than the tariff.

We begin with a situation of free goods trade, and no initial capital relocation. Production is represented by P in Figure 1. Country H imports DC of cotton and exports DP of steel. Now country H levies a tariff on all goods imports of 100τ percent, and a tax on the earnings of capital placed in the foreign country of $100t$ percent. All foreign earnings are repatriated.

If $t/\tau > \alpha - 1$, no capital relocation occurs. The tariff raises the price of cotton in country H by 100τ percent, but the difference in nominal returns to capital between the two countries is less than $100t$ percent.

If $t/\tau < \alpha - 1$, capital relocation occurs. The price of cotton in H rises by 100τ percent. By the Rybczynski theorem, because of the capital relocation, production of cotton rises and that of steel falls in country H. The opposite occurs in country F. The excess demand for cotton to be imported into country H falls. Country H now pays for its imports of cotton partly through its earnings on foreign capital, partly through its exports of steel. As capital continues to relocate, exports of steel continue to fall. Eventually, exports of steel are zero, and the value of cotton imports is exactly equal to the value of earnings on foreign placed capital. At that point, we have the price of cotton 100τ percent higher in country H than in country F and the price of steel equal in the two countries (because country H was exporting steel). But there still is an incentive for more capital to move to country F (i.e., the capital market is still not in equilibrium). As this relocation continues the production of steel falls further in country H, thus giving rise to a demand to *import* steel into country H; i.e., the price of steel begins to rise. However, steel will not be imported into country H until the price of steel there rises by 100τ percent because there is a tariff on all imports, steel as well as cotton.

As the price of steel rises in country H, the relative price of steel in terms of cotton rises too. This raises the real and also the nominal rental to capital in country H. This reduces the incentive for capital to move to country F. As soon as the price of steel rises sufficiently to make the relative prices in the two countries differ only by $(t+\tau)/\alpha < \tau$, the capital relocation ceases. At that point, the net nominal returns to home-owned capital are equal in the two countries. No more capital leaves country H. We have $(t+\tau)/\alpha < \tau$ because we began with a situation where $t/\tau < \alpha - 1$.)

We can see how the pattern of trade has changed. Country H now imports cotton, but pays for it through capital exports, not steel exports. The nominal prices of cotton differ by 100τ percent between the two countries. The nominal prices of steel differ by *less* than 100τ percent, the price of steel being higher in country H. Nevertheless, steel is not imported into country H because of the tariff on all imports.

Actually, the tariff and tax incentives create a situation analogous to that of capital immobility—except that now *steel* is in a sense immobile (not by assumption, but by market forces) instead of capital.

Before we present a diagrammatic illustration of the equilibrium after factor relocation in the presence of tariffs and taxes, we can derive some more algebraic results.

We can analyze the impact of *differential* tariffs on cotton imports and steel imports. Let these be represented by τ_C and τ_S. Let $100t$ percent again be the tax on foreign earnings. We can demonstrate a relationship between τ_C, τ_S and t which will determine, under conditions of capital relocation, whether country H, in final equilibrium, will be importing cotton or steel and which good will disappear from international trade; i.e., we can show the conditions necessary for pattern 2 or pattern 3 type of exchange to occur (see Section I).

We saw above, that for $\tau_C=\tau_S>0$, $t>0$, the equilibrium differential in relative prices was $dp/p=-(t+\tau)/\alpha$. Now, $dp/p=dP_S/P_S-dP_C/P_C=-(t+\tau)/\alpha$. For $\tau_S=\tau$, we had $dP_C/P_C=\tau$.

$$\therefore\ \frac{dP_S}{P_S}=\frac{dp}{p}+\frac{dP_C}{P_C}=-\left(\frac{t+\tau}{\alpha}\right)+\tau$$
$$=\frac{\tau(\alpha-1)-t}{\alpha}$$

But, capital relocation occurs only if $t/\tau<\alpha-1$, i.e. $\tau(\alpha-1)>t$. Therefore, $dP_S/P_S=[\tau(\alpha-1)-t]/\alpha>0$; also, $dP_S/P_S=\tau-\tau/\alpha-t/\tau=\tau-(\tau+t)/\alpha<\tau$, so $\tau>dP_S/P_S>0$.

Thus, if $\tau_S=\tau_C=\tau$, steel is not imported into the home country, because the domestic price of steel is less than $100\tau_S$ percent greater than the foreign price. Steel would be imported if τ_S were lowered to the point where τ_S was less than the difference between home and foreign prices, i.e., if $\tau_S<[\tau_C(\alpha-1)-t]/\alpha$.

Thus, for a situation where capital relocation occurs steel will be the only good imported into country H if

$$\tau_s<[\tau_C(\alpha-1)-t]/\alpha$$

What this means in terms of the capital relocation process described above is as follows. Initially, country H imports cotton and exports steel. After the tariffs and tax are levied, capital relocates to country F. Eventually, the point is reached, because of the Rybczynski effects, where no steel is exported. Cotton imports are paid for totally out of earnings on foreign-placed capital. But the capital relocation continues beyond that point, because the capital market is not yet in equilibrium, giving rise to an excess demand for steel. The price of steel rises in country H. It can rise no higher than $(1+\tau_S)P_S^F$. When P_S^H reaches $(1+\tau_S)P_S^F$, steel starts being imported into country H, as well as cotton—both imports are paid for out of capital's foreign earnings. But, the capital relocation does not cease, because relative prices differ between the countries by an amount sufficient to retain the incentive for capital relocation.

Again, by the Rybczynski theorem, the capital relocation causes an increase in production of cotton in country H. This means that as capital moves to country F, desired cotton imports fall. Eventually these desired cotton imports fall to zero, domestic production of cotton in H having expanded sufficiently to satisfy domestic demand. But capital continues to relocate beyond the point where cotton imports are zero, because relative prices haven't changed. As the capital continues to move, cotton production increases in country H. This causes an excess supply of cotton, which lowers the price of cotton to $P_C^H<(1+\tau_C)P_C^F$. When P_C^H falls low enough so that the relative price differential between the countries has narrowed so as to cut off the incentive for capital relocation, capital relocation ceases. And, at that final equilibrium, steel is the only good being imported into country H.

We can now summarize our algebraic results:

A capital relocation will occur if $t/\tau_C<\alpha-1$.

If $t/\tau_C<\alpha-1$, cotton will be the only good imported into country H if $\tau_S>[\tau_C(\alpha-1)-t]/\alpha$, and the equilibrium production point will be "in the vicinity" of J in Figure 1. Furthermore, we will have $dp^*/p=-(t+\tau_C)/\alpha>-\tau_C$: also, $\tau_s>dP_S/P_S=[\tau_C(\alpha-1)-t]/\alpha>0$.

If $t/\tau_C<\alpha-1$, and $\tau_S<[\tau_C(\alpha-1)-t]/\alpha$, steel will be the only good imported into country H, and the equilibrium production point will be "in the vicinity" of N in Figure 1. Furthermore, we will have $dp/p^* = -[\tau_S + t]/(\alpha - 1)$; also, $dP_C/P_C = [t + \alpha\tau_S]/(\alpha - 1) < \tau_C$.

Notice four more particular results:

If $\tau_C = t$, a capital relocation occurs only if $\alpha > 2$.

If there is an undifferentiated tariff as between cotton and steel (i.e., $\tau_S = \tau_C$), and if $t/\tau_C < \alpha - 1$, cotton will be the only imported good, because $\tau_S > [\tau_C(\alpha - 1) - t]/\alpha$.

If $\tau_S = 0$, and $t/\tau_C < \alpha - 1$, steel will be the only imported good because $\tau_S < [\tau_C(\alpha - 1) - t]/\alpha$.

Even if $t = 0$, $\tau_S > 0$, $\tau_C > 0$, capital relocation will occur but the real rental on capital will still fall in country H.

Geometric Presentation

We must now show what the equilibrium looks like, diagrammatically, for a combination of τ and t such as is represented by V in Figure 2. With a tariff τ_v (uniformly applied to steel and cotton imports) and a tax on foreign earnings t_v, we know from our previous analysis that the final equilibrium price differential dp/p between the two countries will be $dp^*/dp = (t_v + \tau_v)/\alpha = W$.

In Figure 3, we show the equilibrium for country H; P is the production point during free goods trade; C is the consumption point. (We hold the foreign price ratio constant, for ease of presentation.)

If we were to consider the case of a tariff of $W = -dp^*/dp = (t_v + \tau_v)/\alpha$, with capital immobility, we would end up with production at P' where the slope of the transformation curve at P' is $(1+W)$ times the slope at P; consumption would be at C'; EE, which runs through P' is parallel to PC, and its slope is equal to the foreign price ratio. At the point C' a domestic community indifference curve crosses EE and has a slope of $(1+W)$ times the slope of EE.

The solution in our capital relocation case is very similar. We know that the domestic price ratio will differ by W from the foreign price ratio. So, through P' we draw the Rybczynski line $R'R'$. Domestic production will be on $R'R'$, southeast of P'. Each point on $R'R'$ represents some quantity of capital outflow; that capital will earn returns abroad, which will be taken home in the form of cotton. These returns must be added to domestic production to get domestic production plus foreign cotton earnings. At C', where EE cuts an indifference curve with a slope equal to $(1+W)$ times the foreign price ratio, we get the equilibrium consumption point; to find the production point, we draw a horizontal line through C', which cuts $R'R'$ at Y. The production point Y is on a new contracted transformation curve $T'T'$.

If there were differential tariffs set (τ_S, τ_C), such that $\tau_S < [\tau_C(\alpha - 1) - t]/\alpha$, so that capital relocation occurred and only steel

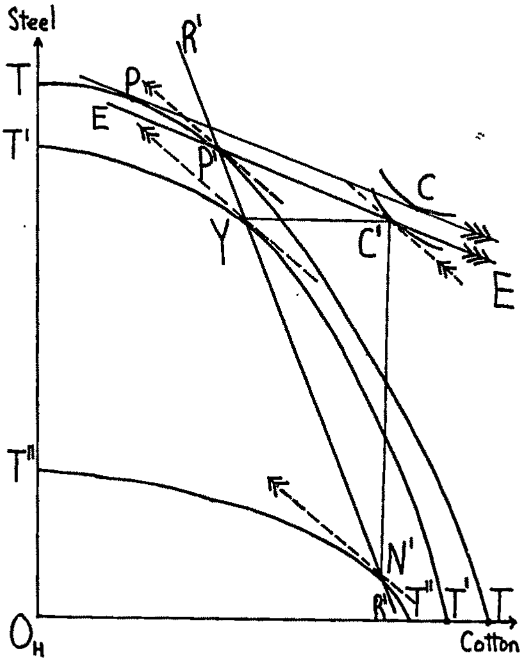


FIGURE 3—EQUILIBRIUM WITH TARIFFS AND TAXES

were imported into country H, equilibrium would be at N' in Figure 3, i.e., at a point on $R'R'$ vertically below C'. It is possible that point N' could not be reached, i.e., that country H might end up specializing in cotton. Point Y can always be reached, i.e., there is no chance of specialization occurring if equilibrium is to be at Y.

That the EE line should have the exact same slope as the foreign price ratio is not at all random. It is dictated by the technology. It is a reflection of the duality of simple general equilibrium models, of the intimate relationship between the Rybczynski and Stolper-Samuelson theorems.

The algebraic proof of this is somewhat arduous. It can be shown intuitively as follows. $P'C'$ is the home country's budget constraint, before capital relocation, expressed in terms of foreign prices. The capital relocation merely causes identically opposite production changes in both countries, leaving the foreign valuation (i.e., in terms of foreign prices) of domestic income unchanged.

Thus, any position of final consumption (utility) achieved through restrictions on barter-goods trade can be achieved through restrictions on factor-relocation trade. (Furthermore, government revenue is exactly the same in the two cases.) The "substitutability" between the two continues to hold. The introduction of factor mobility does not increase the optimization potential a nation has when it seeks to interfere with barter goods trade under conditions of capital immobility. What capital mobility does is provide a potential method of frustrating the intended results of attempted interferences in the barter trade flows.

REFERENCES

C. Iversen, *Aspects of the Theory of International Capital Movements*, Boston 1935, ch. 2.

R. W. Jones, "The Structure of Simple General Equilibrium Models," *J. Polit. Econ.*, Dec. 1965, *73*, 557–72.

———, "International Capital Movements and the Theory of Tariffs and Trade," *Quart. J. Econ.*, Feb. 1967, *81*, 1–39.

M. Longfield, *Lectures on Political Economy*, Dublin 1834.

J. E. Meade, *Trade and Welfare*, London 1955.

R. A. Mundell, "International Trade and Factor Mobility," *Amer. Econ. Rev.*, June 1957, *47*, 321–35.

B. Ohlin, *Interregional and International Trade*, Cambridge, Mass. 1967.

T. M. Rybczynski, "Factor Endowment and Relative Commodity Prices," *Economics*, Nov. 1955, *22*, 336–41.

P. A. Samuelson, "Prices of Factors and Goods in General Equilibrium," *Rev. Econ. Stud.*, 1953, *21*, no. 1, 1–20.

W. R. Stolper and P. A. Samuelson, "Protection and Real Wages," *Rev. Econ. Stud.* Nov. 1941, *9*, 58–73.

J. H. Williams, "The Theory of International Trade Reconsidered," *Econ. J.*, June 1929, *39*, 195–209.

Peasants, Procreation, and Pensions

By PHILIP A. NEHER*

The neoclassical theory of population employs cost-benefit analysis to predict the size of family which fertility decision makers desire. While the theory has not been articulated in detail, its broad outlines have been made clear enough to guide the formulation of testable hypotheses.[1] Roughly put, the theory treats children as producers' and consumers' durable goods, yielding a stream of benefits which are compared with a stream of costs. Births occur if the present value of the benefits exceeds the present value of the costs.

The theory has been used by Malthusians to rationalize the casual observation that population growth rates increase during the early stages of economic development and then, perhaps, fall later on as even higher levels of per capita income are attained. The implications of this behavior have been explored in dynamic as well as static macroeconomic models.[2]

As useful as these models have been in improving our understanding of growth and development processes, it must be said that the underlying Malthusian theory of population lacks precisely formulated micro-foundations. These foundations are important for familiar reasons. First, they are required to sharpen the predictive accuracy of the theory itself. Second, they must be secured before effective population control schemes can be devised.

Until recently, the population control effort has focused on devising inexpensive and morally acceptable birth control measures. Yet the demographic history of France and Ireland suggests that these measures are not necessary conditions for population control.[3] They are probably not sufficient conditions either. People must *want* to reduce *desired* family size. Modern birth control measures simply make it easy to do.[4]

If parents believe they can make themselves better off by having large families, they will do so. If economic considerations impinge on fertility decisions, public policy measures designed to lower fertility rates are likely to fail unless they strike directly at basic economic motives for having children. Thus, propaganda and pills alone are unlikely to have a major impact on reducing fertility if it is clear to family planners that they can improve their condition by having more children.

A major motive for having children in primitive societies is the *pension motive.* Parents invest in their children by bearing their rearing costs in anticipation of retirement when their children, in turn, will support them. I supress other motives for having children and thereby sacrifice a

---

* Associate professor of economics, University of British Columbia. I am grateful for discussions with Keizo Nagatani, John Cragg, and Anthony Scott, and for comments by an anonymous referee.

[1] A widely available statement of the theory is in Milton Friedman (pp. 207–11). Harvey Leibenstein (pp. 159–70), outlines the theory in the context of economic development. Examples of empirical applications are found in Gary Becker and Theodore Shultz.

[2] John Buttrick (1958) has made population growth endogenous in Robert Solow's model. Richard Nelson used a Malthusian approach in formulating his well known "low-level equilibrium trap." Jürg Niehans extends Nelson's model to allow for capital accumulation. John Conlisk explores a fascinating model which is related to Niehan's. Buttrick's (1960) contribution is similar to Nelson's and is an excellent exposition of the Malthusian model. Eric Davis and Robert Merton have explored the implications of Malthusian population growth for the optimal time path of saving in a growing economy.

[3] For other examples, see E. A. Wrigley.

[4] Leibeinstein is a major proponent of this view (see p. 159).

large measure of realism in an effort to gain precision. I shall make a number of simplifying assumptions in addition as I go along—some of them are quite drastic. What emerges is an economic theory of population in a primitive economy which, however abstract and unrealistic, seems to illuminate fundamental dimensions of human fertility in societies where children commonly provide for their parents during their retirement years. The model is thoroughly neoclassical; no new concepts are involved. I am trying to forge a new tool of analysis, but with off-the-shelf components.

The pension motive for having children will be examined in the context of a primitive economy and a natural life cycle of dependent childhood, productive parenthood, and dependent grandparenthood. The life cycle gives rise to an optimal stock of children from the parents' point of view and they control their fertility accordingly. The optimal stock of children is associated with an optimal lifetime consumption profile which is calculated on the assumption that future births are exogenous events.

I shall be analyzing a primitive economy where the only property is land, and the rights to it are vested in families or extended families. People do not, as individuals, have a claim on property income, nor do they have a claim on the fruits of their own labor. Income is distributed in accordance with a *share alike ethic* whereby all members of the family have equal claim to the product whether they work or not.

I assume further that goods cannot be carried forward through time. Parents cannot provide for their unproductive retirement years by storing goods. Thus, the *only* way for parents to consume while retired is by means of an income transfer from the working generation. This transfer is assured by the share alike ethic which serves as an unwritten and continuing intergeneration contract whereby the workers agree to support the retired.

If fertility were exogenous, and a constant force, there would be no decisions to make. Life cycles would simply repeat themselves as exact replicas of one another. But fertility *is* a decision variable in the model which I shall set out in the next section. People *can* make themselves, and future generations, better or worse off, by controlling their fertility. What *will* they do?

How do egoism and shortsightedness bear on the fertility decision?

What is the consequent optimum size of family, in equilibrium, from the point of view of the independent decision maker? Is there "overpopulation" in some sense, and is it significant due to the pension motive alone?

How might feasible institutional reform impinge upon the fertility decision? Would the introduction of alternate sources of pensions significantly reduce the pension motive for fertility?

These questions are explored in subsequent sections, followed by some highly tentative conclusions.

I. The Model

Imagine a continuing (extended) family unit composed of three equally spaced generations: grandparents (G) who have retired and perform no economic functions; parents (F) who work a fixed amount of land;[5] and children (S) who, like the grandparents, are dependent on the working parents. The total population (P) equals $S+F+G$. The passage of a unit of time turns children into parents, parents into grandparents, and grandparents into their graves. Simultaneously, a new generation of children appears.

I assume that the fertility decision is made and effected by the parents at the instant they move into the parent generation. The parents thus look forward to one full period during which they work and their children are dependent, and one full period during which their children work and they (i.e., the parents) are dependent. These two periods are followed by a third during which the children of the original parents are dependent grandparents.

---

[5] One can think of land as a composite, non-augmentable non-labor input.

The consequences of a fertility decision thus span three periods, but the parents are dead during the third. Nevertheless, parents may have some concern for the state of the world after they are dead, while their children are still alive. To allow for this possibility, I designate a utility function which embraces the three periods during which the children will live. The parents' egoistic utility depends upon their own consumptions (and, because of the share alike constraint, everyone else's too) during the first two periods. The parent's concern for the future is represented by the utility of consumption enjoyed by their children (and everyone else) during the third period. I assume that utility is additive, so the parents choose to have the number of children (S_1) which will maximize[6] the following utility function, where D and E represent discount factors:

$$(1) \qquad U = u(c_1) + D\cdot u(c_2) + E\cdot u(c_3); \qquad u'(c) = \mu > 0; \quad u''(c) = \mu' < 0$$

The constraints are production conditions, the share alike ethic, the fertility of future generations and interperiod mortality.

The production relation, $X=f(F)$, has the traditional form with first increasing and then decreasing marginal returns to labor as more labor is added to the land. It is illustrated in Figure 1. Note that the marginal product of labor (w) equals the average product of labor (X/F) where labor's average product is a maximum.

The share alike ethic has the obvious implication that the consumption effects of fertility behavior are felt equally by everybody, not just by the parents.

The fertility of future generations is in the hands of future parents and exogenous to the current period decision makers.[7]

Interperiod mortality is represented by p, the probability that a child (S) will live long enough to become a parent (F) and by q, the probability that a parent (F) will live long enough to become a grandparent (G). For example, $F_2 = pS_1$ and $G_3 = qF_2 = pqS_1$, where the subscripts refer to periods.

These constraints are summarized in equations (2) through (4). Fixed and exogenous variables are denoted by a superscript bar.

$$(2) \qquad c_1 = \frac{\bar{X}_1}{P_1} = \frac{f(\bar{F}_1)}{S_1 + \bar{F}_1 + \bar{G}_1}$$

$$(3) \qquad c_2 = \frac{X_2}{P_2} = \frac{f(pS_1)}{\bar{S}_2 + pS_1 + q\bar{F}_1}$$

$$(4) \qquad c_3 = \frac{\bar{X}_3}{P_3} = \frac{f(p\bar{S}_2)}{\bar{S}_3 + p\bar{S}_2 + pqS_1}$$

The share alike ethic compels the first period product ($\bar{X}_1$), produced by the existing parents ($\bar{F}_1$'s), to be shared alike with the retired grandparents ($\bar{G}_1$'s) and however many children (S_1's) are chosen. First-period consumption (c_1) falls as S_1 rises.

$$(5) \qquad \frac{dc_1}{dS_1} = -\frac{1}{P_1}c_1$$

---

[6] To preserve simplicity, I assume that S_1 is a continuous and unbounded control variable. The skeptical reader can set up the problem with S_1 as a discontinuous and bounded control. The results are substantially the same.

[7] There seems no very good way to make future fertility endogenous. One possibility is the game-theoretic approach. See Edmund Phelps and R. A. Pollak. But that is another exercise.

The first-period consumption consequences of fertility are illustrated by the dotted c_1 line in Figure 1. Starting with a point like M (ignore the c^* line for now), maximum c_1 is obtained by having no children ($S_1=0$); having more reduces per capita product.

Second-period per capita consumption increases as more workers are added so long as labor's marginal product, w, exceeds per capita consumption, $c=X/P$, (not per worker product, X/F).

$$\frac{dc_2}{dS_1}=\frac{p}{P_2}(w_2-c_2) \tag{6}$$

Starting at point M, second-period consumption, denoted by the dotted c_2 line, is zero if $S_1=0$. It rises with increased fertility up to point N where the marginal contribution of an additional worker equals average consumption ($w_2=c_2$), and then falls.

Third-period per capita consumption falls with larger numbers of first-period children. A given amount of third-period product yields less per capita consumption if it must be shared with a larger retired generation.

$$\frac{dc_3}{dS_1}=-\frac{pq}{P_3}c_3 \tag{7}$$

A necessary condition for a maximum U with respect to S_1 is that dU/dS_1 vanish.

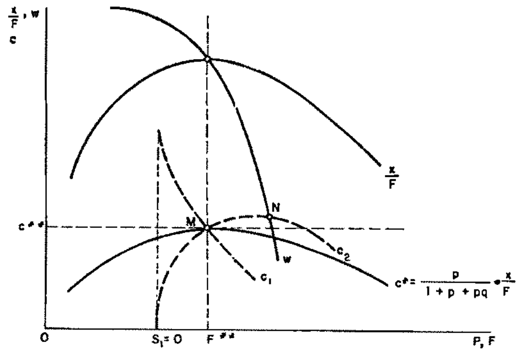


FIGURE 1

Using (5), (6) and (7), along with (1), the condition is

$$\hat{w}_2=\frac{1}{pD}\left(\frac{\mu_1}{\mu_2}\cdot\frac{P_2}{P_1}\hat{c}_1+Dp\hat{c}_2+E\frac{\mu_3}{\mu_2}\cdot\frac{P_2}{P_3}pq\hat{c}_3\right) \tag{8}$$

where the superscript hat denotes a value associated with utility maximizing by the parents. Equation (8) is the condition for *economic* equilibrium.

The expression appears less formidable when applied to a steady state, or Golden Age,[8] marked by a stable and stationary population.[9] These conditions imply constant product, steady per capita incomes (and utilities) and an unchanging marginal product of labor (w). For a Golden Age, (8) is written

$$\hat{w}^*=\frac{1}{pD}(1+pD+pqE)\hat{c}^* \tag{8$'$}$$

where the asterisk denotes a Golden Age value.

This condition is still not readily interpreted. However, I note that consumption per capita is related to output per worker in a Golden Age by the equilibrium demographic structure of the population. Specifically, since $F^*/P^*=p/(1+p+pq)$, I can write

$$c^*=\frac{p}{1+p+pq}\cdot\left(\frac{X}{F}\right)^* \tag{9}$$

This is the c^* line in Figure 1. Note that c^* reaches a maximum value where $w=X/F$.

Using (9) and (8′), I obtain an expression relating output per worker to labor's marginal product

---

[8] A Golden Age is identified in modern growth literature as a balanced growth state where every variable, if it grows at all, grows at the same proportional rate. Here, a Golden Age refers to a constant population with each generation proportional in size to every other.

[9] Demographers call a population stable if its demographic structure is unchanging. A stationary population has constant numbers. See G. W. Barcley.

$$\hat{w}^* = J \cdot \left(\frac{\hat{X}}{F}\right)^* \tag{10}$$

$$J = \frac{1}{D} \cdot \frac{1 + pD + pqE}{1 + p + pq}$$

$$J_D < 0, \qquad J_E > 0$$

This equation combines conditions for economic equilibrium (the parents are satisfied with the size of the oncoming generation), and demographic equilibrium (there is a Golden Age where the population is stable and stationary).

II. The Golden Rule of Fertility[10]

Equation (10) relates the Golden Age values of w and X/F. If $J=1$, the Golden Age population enjoys the maximum sustainable level of per capita income at point M in Figure 1. There, $c^*=c^{**}$, $F^*=F^{**}$ and $P^{**}=(1+p+pq)F^{**}$. The double-star denotes a *Golden Rule* value. Each generation of parents can say "we are procreating for future generations as we would have had past generations procreate for us."

The value of J depends upon interperiod survival rates (p and q) and the discount factors assigned to future utilities (D and E). The value of J will equal unity if

$$D = \frac{1 + pqE}{1 + pq} \tag{11}$$

Then the Golden Rule is being observed. Clearly, a sufficient condition for compliance is that $D=E=1$. This means that parents weigh equally their utilities while working and retired *and* that they have an equal regard for their children's utility. Such farsightedness ($D=1$) and altruism ($E=1$) harmonizes the self-interest of individual parents and of the continuing family.[11]

Curiously, generalized discounting of the future ($E<D<1$) can be consistent with the Golden Rule. For example, if $p=q=E=1/2$, then $J=1$ if $D=9/10$. However, there is no uniform, non-zero, rate of discount (r) consistent with the Golden Rule. If $D=1/(1+r)$ and $E=1(1+r)^2$, (11) holds only if $r=0$.

But suppose parents care not at all for their children apart from sharing alike with them while alive so that E equals zero. Suppose further that D equals one. Then

$$J = \frac{1 + p}{1 + p + pq} < 1$$

and the corresponding Golden Age will be marked by "overpopulation" with labor's marginal product less than its average product. But the degree of overpopulation may or may not be "large," depending on the nature of the production function and the magnitude of intergeneration mortality.[12]

I have been working with a utility function in which the quality of life (consumption per capita) alone counts. But it may be that children yield a flow of consumption delights for their parents and are therefore valued for themselves.[13] In that case, the maximizing behavior of parents results in overpopulation in the sense I have used it.

---

[10] Compare Phelps' "Golden Rule of Procreation." The concepts are related but derived in different contexts. Phelps' Rule is for a dynamic Golden Age with capital accumulation and children valued, in part, as consumption good. My Rule is for a static Golden Age where children are valued solely as an investment for pension purposes.

[11] The analogue in growth theory is well known. Altruism and farsightedness will lead private savers toward a "consumption turnpike" along which consumption per head is the sustainable maximum. See Samuelson (1965).

[12] For example, if $p=q=1/2$ then $J=6/7$. If the production function is of the form $X=L^2/(A+B\cdot L^3)$, then $L^{**}=(A/2B)^{1/3}$ and $(X/L)^{**}=(2/3)(A/2B)^{1/3}$. But if $J=6/7$, then $L^*=(8A/13B)^{1/3}$ and $X/L^*=(13/21)(8A/13B)^{1/3}$. The ratio of the latter X/L to the former is 0.995. I would judge overpopulation to be "small" in this case.

[13] This is a common assumption in the literature. See, for example, Becker (p. 211) and Phelps (p. 181).

It is left to the reader to insert children as an argument in the utility function and work out the consequences.

III. A Fundamental Market Failure

It is well known that economic inefficiency is likely to occur if the costs and benefits relating to a decision are not wholely internalized to the decision making unit. It was found in the preceding section that the Golden Rule will be violated if parents are farsighted with respect to their retirement years ($D=1$) but neglect the welfare of those who live on after the parents die ($E=0$) but who, nevertheless, are affected by the parents' fertility.

Except in societies which practice instant reincarnation,[14] I suspect that people have little regard for the world subsequent to their own deaths so that the special case where $D=1$ and $E=0$ particularly merits attention. To simplify matters even further, I set $p=q=1$ (no interperiod mortality).

The intemporal consumption opportunities were derived in Section I and appear in Figure 1 as the dotted c_1 and c_2 lines. The initial point is M which is now interpreted as a Golden Rule point. Will the Rule be violated or will the current generation of parents choose to just replace themselves?

The consumption information in Figure 1 is translated into Figure 2 where it appears as the dotted consumption opportunities line (labelled cc). Starting from the Golden Rule consumption point at M, the current generation of F's can get more c_1 at the expense of c_2 up to the point where $c_1 = \overline{X}_1/(\overline{F}_1+\overline{G}_1)$. Moving the other way, they can get more c_2 at the expense of c_1 up to the point where $w_2=c_2$.

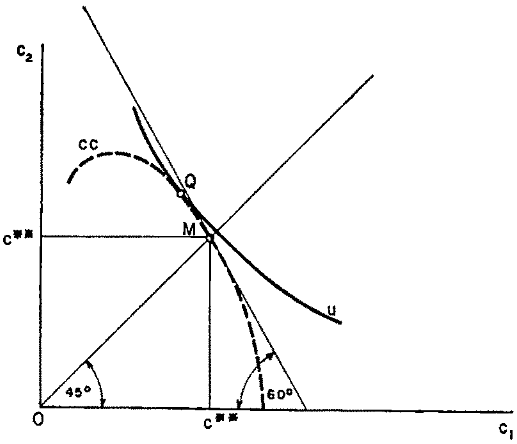


FIGURE 2

The current generation of parents will choose to just replace themselves, staying on the Golden Rule path, only if $c_1=c_2=c^{**}$ maximizes the two-period utility function, equation (1) with $E=0$ and $D=1$. This function can be drawn as a set of intertemporal indifference curves, symmetrical about a 45° ray out of the origin with slopes

$$\frac{dc_2}{dc_1} = -\frac{\mu_1}{\mu_2}$$

At the Golden Rule point, M, $c_1=c_2$ so that $\mu_1=\mu_2$ and

$$\frac{dc_2}{dc_1} = -1$$

The Rule will be violated unless the slope of $cc=1$ at M.

The slope of cc at M is found by dividing (6) by (5), observing that $P_1=P_2$ in a Golden Age and setting $p=1$.

$$\frac{dc_2}{dc_1} = -\frac{w^* - c^*}{c^*}$$

Per capita consumption and per worker product are related by (9). With $p=q=1$

$$c^* = \frac{1}{3}\cdot\left(\frac{X}{F}\right)^*$$

If, in addition, the Golden Rule is being observed, $(X/F)^{**}=w^{**}$ so that $w^{**}=3\cdot c^{**}$ and

---

[14] By instant reincarnation, I mean a return to the world as an S just after having been a G.

$$\frac{dc_2}{dc_1} = \frac{3c^{**} - c^{**}}{c^{**}} = -2$$

is the slope of cc at M.

I conclude that the Golden Rule will be violated. At point M, technology and the share alike ethic permit parents to trade off one unit of consumption in the first period for two in the second. But they are prepared to trade even. Second-period consumption appears relatively cheap to them and they will opt for more of it by procreating in excess of Golden Rule standards, driving c_1 below c^{**} and c_2 above it, moving along cc to point Q. The family moves toward an "overpopulated" condition.

I shall not trace out the adjustment path. That is a very difficult task, even in this simple model.[15] However, the eventual equilibrium must satisfy (10) which, with $p = q = D = 1$ and $E = 0$, reads

$$w^* = \frac{2}{3} \cdot \left(\frac{X}{F}\right)^*$$

Per capita consumption falls short of its maximum sustainable (Golden Rule) value.

Why does this overpopulation occur? After all, the parents are constrained by the share alike ethic. Maximizing *their own* utility must have a salutory effect on *everyone else's* utility. Or does it? The trouble is simply the fact of human mortality and that people tend to discount the state of the world subsequent to their deaths. In this case, having children imposes costs on the parents when they are parents, generates benefits when the parents have become grandparents, *and* imposes costs once again *after the original parents are dead* and their children have become dependent grandparents. The private fertility decision makers (parents) are able to externalize some of the costs of having children to future generations by dying. The private cost of having children falls short of the social cost, if we can think of the continuing family unit as the "society." This illustrates once again the *fundamental market failure* which exists when the consequences of current decisions carry beyond the decision makers' planning horizon. In this case, the welfare of the children when retired, of their children, and of their children's children are not taken into account.[16]

IV. Private Property and Other Social Contrivances

If the pension motive for having children is strong, one wonders how the desired size of family would change if alternate sources of pensions were available.

Share alike distribution of family product is an obvious method for ensuring that the retired have income during their retirement years. A major rationale for the share alike ethic may be precisely because it leads to automatic pensions in a world where interperiod storage of goods is impossible and financial markets are not developed. But suppose that institutions of private property and market contract spring up to link together contemporary families and to provide intergeneration contracting through marketing of contracts for delivery of goods in the future (pensions). Highly developed institutional forms are not required to perform these functions. A changeover to vesting the ownership of land in *individuals* rather than continuing family units will do the trick. The retired gradually sell off titles to their land to working generations in return for goods. The working generations have an incentive to buy the land, for they look forward to trading their land for goods during their own retirement years. The current purchase of land with goods implies the future delivery of goods for the working generation when they retire.

---

[15] The adjustment path is described by a third (at least) order difference equation. The path is convergent for specifications of the adjustment process which put realistic upper and lower bounds on the control variable S_1. Otherwise, the population is likely not to converge.

[16] Note that this occurs independent of "like father, like son" effects. This model takes the fertility of future generations as exogenous.

Money is an alternate social contrivance. Workers buy money from the retired in exchange for goods in anticipation of selling the money for goods while retired. The current purchase of money with goods implies the future delivery of goods for the working generation when they retire.

The impact of these alternate pension schemes on desired family size depends critically upon the extension of the market beyond the family unit. Even if the share alike ethic has withered away, the pension motive for procreation will remain if the "market" for money or land is internal to the family. In that case, parents will perceive the possibility of improving the terms on which they can trade land (or money) for goods to their children, if they have more children (at least up to a point).

The pension motive disappears, however, if markets extend beyond the bounds of the family unit and if the share alike ethic is abandoned. Then having children imposes rearing costs but confers no benefits on the parents. Some other motive would have to account for a continuing population.

Traditional ways are not easily given up, however, and new institutional forms are not likely to be adopted unless it appears advantageous to do so. In particular, the social contrivances I have mentioned, private property and money, will not spring into existence spontaneously. And even though there may be a collective will to reduce population size, and even though these contrivances may serve that end, parents may persist in providing their pensions in traditional ways if their children are good investments compared with available alternatives. How do children compare as investments with land or money?

Consider the simplest case: no intergeneration mortality, no discounting of retirement utilities and no weight given to the future after death. What is the two-period consumption rate of return in equilibrium on land and money?

These contrivances give rise to a transfer of goods from working to retired generations which simulates a costless carry forward of goods. The workers can trade one for one, consumption now for consumption in the future.

But note that the two-period consumption rate of return on children is also zero, under equilibrium conditions in the same model. This is clear from the $c_1 = c_2$ property in any Golden Age and the assumed lack of time preference.

I conclude that the individual fertility decision makers will perceive no clear advantage from abandoning the use of their children for pension purposes, if the alternatives are private property or the use of money, and the share alike ethic will survive.

Perhaps other alternatives would be more attractive. Suppose, for example, that money bore a positive rate of interest. By money, I mean assets issued by agents outside the primitive society: interest bearing liabilities of public or private financial intermediaries. I shall call these liabilities "bonds." An example would be deposit accounts in cooperative credit unions which mobilize saving in a primitive rural sector for investment in a modern industrial sector.

If bonds were available to a share alike society, would parents desire to reduce their family size? The answer is yes—the good asset (bonds) drives out the bad asset (children). An initial Golden Age, before the introduction of bonds, is depicted at point M in Figure 3. The return on children is zero reflecting time preference.

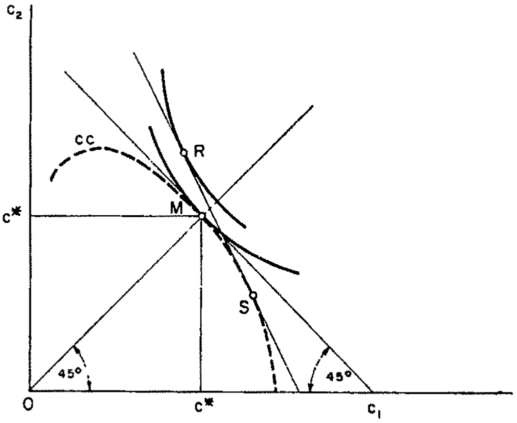


FIGURE 3

Now introduce bonds which bear a positive interest rate. Point *M* is no longer consistent with utility maximizing, for parents will perceive the possibility of generating investable funds by having fewer children and trading those funds forward at a positive rate of interest. The decline in fertility is denoted by a movement along *cc* from *M* to *S*. The corresponding c_1 and c_2 are consumptions that would be enjoyed if bonds were not purchased. But now the c_1 in excess of c^* can be traded forward by buying bonds and a new consumption point achieved at *R*, with $c_2 > c_1$ and the return on children equalized with the return on bonds.

This is the reaction of the first generation of parents subsequent to the introduction of bonds: fertility is reduced below Golden Age levels. I infer that subsequent generations will not be satisfied to procreate at levels which would maintain population size. The population will wither away unless there are other motives for having children.

V. Conclusions

Few economists or policy makers would take the view that no important country today is overpopulated in the Golden Rule sense discussed in Section II. Most of us would argue that individual fertility decisions have given too much weight to the quantity of life relative to its quality. I am a professed naive optimalist. I would go all the way to favor the quality of life alone, and opt for the Golden Rule.

My analysis in Section III suggested a fundamental market failure which is rooted in human mortality and a chronic inability of most of us to care much about the state of the world subsequent to our deaths. The effects of the market failure on population size are offset by positive time discounting which assigns a lower weight to utility while old and retired than to utility while young, working, and raising children. The net effect could lead to Golden Rule fertility. But if parents are optimistic of living into retirement, assign relatively high weight to their retirement utility, and care not at all for future generations, the model predicts an equilibrium population in excess of Golden Rule levels on account of the pension motive *alone*.

A more complete model would take uncertainty into account: the uncertainty of survival to retirement and the uncertainty that the children will live on to support their parents in retirement. Risk averting parents would "overprocreate" to reduce the chances of starvation in old age. This effect is attenuated if extended family arrangements or some other social insurance scheme is in effect to take care of the childless retired.

But even the simple certainty model I have worked with goes far enough to demonstrate that one should not rule out the possibility of Golden Age populations in excess of Golden Rule levels being generated by the pension motive *alone*.

This suggests that population size can be significantly reduced if alternate pension schemes are made available. However, they must appear superior to children or they will not be adopted by parents. The analysis in Section IV showed that a positive rate of return on bonds will eventually eliminate the pension motive (and the model population) if parents give equal weight to their utilities in their two remaining life cycle periods. It is easy to show more generally that bonds must carry a two-period rate of return in excess of the parent's two-period rate of discount in order to have the fertility attenuating effect.

The implications of these observations are two-fold:

> They may help explain lower levels of fertility in urban areas where parents have easy access to highly developed financial markets for pension purposes and, perhaps for that reason, family ties are loose and life is impersonal.[17]
>
> If a population control problem is thought to exist, a control scheme might well include the provision of pensions by means of private or social pension plans. A combination of pensions and pills might well be more effective than pills alone.

REFERENCES

G. **Barcley,** *Techniques of Population Analysis,* New York 1958.

G. **Becker,** "An Economic Analysis of Fertility," in *Demographic and Economic Change in Developed Countries,* Nat. Bur. Econ. Res. Conference report, Princeton 1960.

J. **Buttrick,** "A Note on Growth Theory," *Econ. Develop. Cult. Change,* Oct. 1969, *9,* 75–82.

———, "A Note on Professor Solow's Growth Model," *Quart. J. Econ.,* Nov. 1958, *72,* 633–36.

J. **Conlisk,** "A Modified Neoclassical Growth Model with Endogenous Technological Change," *Southern Econ. J.,* Oct. 1967, *33,* 199–208.

E. **Davis,** "A Modified Golden Rule: The Case with Endogenous Labor Supply," *Amer. Econ. Rev.,* Mar. 1969, *59,* 177–81.

M. **Friedman,** *Price Theory,* Chicago 1962.

H. **Leibenstein,** *Economic Backwardness and Economic Growth,* New York 1960.

R. C. **Merton,** "A Golden Golden-Rule for Welfare-Maximization in an Economy with a Varying Population Growth Rate," *Western Econ. Rev.,* Dec. 1969, *7,* 307–18.

R. R. **Nelson,** "A Theory of the Low-Level Equilibrium Trap in Underdeveloped Countries," *Amer. Econ. Rev.,* Dec. 1956, *46,* 894–908.

J. **Niehans,** "Economic Growth With Two Endogenous Factors," *Quart. J. Econ.,* Aug. 1963, *77,* 349–71.

E. S. **Phelps,** "The Golden Rule of Procreation," in his *Golden Rules of Economic Growth,* New York 1966, 176–83.

——— **and R. A. Pollack,** "On Second Best National Saving and Game-Equilibrium Growth," *Rev. Econ. Stud.,* Apr. 1968, *34,* 185–99.

P. A. **Samuelson,** "A Catenary Turnpike Theorem Involving Consumption and the Golden Rule," *Amer. Econ. Rev.,* June 1965, *55,* 486–96.

T. P. **Shultz,** "An Economic Model of Family Planning and Fertility," *J. Polit. Econ.,* Mar./Apr. 1969, *77,* 153–80.

R. M. **Solow,** "A Contribution to the Theory of Economic Growth," *Quart. J. Econ.,* Nov. 1956, *70,* 537–62.

E. A. **Wrigley,** "Family Limitation in Pre-Industrial England," *Econ. Hist. Rev.,* Apr. 1966, *19,* 82–109.

---

[17] This implication is hardly new. See, for example, Friedman (p. 208).

A Model of Soviet-Type Economic Planning

By MICHAEL MANOVE*

Each year, planning agencies in the Soviet Union construct an annual economic plan for the calendar year that follows. The annual plan is an important element in the Soviet schema for achieving long-term economic growth. Goals of less detailed long-range plans must be considered in the design of the annual plan. Furthermore, the plan is supposed to expose and remedy weak links in the economic structure. But the most important and obvious purpose of the annual plan is to ensure that the economy will function in a reasonable way from day to day.[1]

The material supply plan is a major component of the Soviet-type annual plan. It specifies aggregate output targets and other production indices from which output targets and certain production indices for individual productive units are derived. The traditional procedure for constructing the material supply plan is very complicated and involves a large planning bureaucracy. In fact, it is difficult to determine from the descriptive literature how rational the planning procedure is, or even how different elements of the procedure fit together.[2] In this paper, an attempt is made to establish a theoretical underpinning for this traditional Soviet-type planning procedure.[3] In particular, we try to analyze how, in theory, a planning procedure of this type can produce a consistent plan, i.e., a plan in which the supply and demand for each commodity is balanced.[4] In order to simplify our task, we shall not consider the Soviet planning system itself. Rather, we shall concentrate our analytical efforts on a more transparent, though not dissimilar, planning system—namely, that of the Autonomous Soviet Republic of Morozhenoe. The planning procedures of this republic oblige us by incorporating only the most basic elements of traditional Soviet-type systems.

I. Characteristics of an Annual Plan

Morozhenoe, incorporated as an Autonomous Republic of the Russian Soviet Federated Socialist Republic shortly after World War II, is the only Soviet republic that has been completely autonomous in the economic sphere. Even more surprisingly, despite the fact that Morozhenoe has a smaller gross product than

---

* Assistant professor of economics, University of Michigan. I am indebted to Maria Augustinovics, Michael Bruno, Evsey D. Domar, Richard S. Eckaus, Duncan K. Foley, Martin L. Weitzman, and a referee for their extremely helpful comments and suggestions at various stages of this work. I wish to acknowledge financial support and encouragement from the Comparative Economics Program of the University of Michigan.

[1] According to G. Sorokin (p. 225), the long-term plan outlines ways of achieving the main tasks of economic development, the five-year plan elaborates construction and operational plans in greater detail, and the annual plan itemizes the five-year program and facilitates economic management. The annual plan is legally binding.

[2] The most extensive description by a western scholar of Soviet material-supply planning is contained in the unpublished doctoral dissertation of Herbert Levine (1961). A condensation of the dissertation appears in his 1959 paper. I. A. Evenko (pp. 72–88) is another source of general descriptive material. For a description of the Polish procedure, see J. M. Montias (1962, pp. 6–32, 76–114).

[3] An excellent survey of possible theoretical approaches is contained in Montias (1959).

[4] For general information on the consistency of Soviet plans, see Michael Ellman. Montias (Oct. 1962) discusses the possibility of utilizing the decomposability of the A-matrix to produce consistent plans.

even Darien, Connecticut, this little republic is virtually self-sufficient and engages in little external trade.

Wishing to embark on a program of rapid industrialization, the Morozhenoens decided to construct a "command economy," of which an annual plan of material supply was to be an integral part. Because the system envisioned was to be basically non-market, the annual plan would have to serve as the main guide for the producing units of the economy in their day-to-day operations. The system for constructing such short-term material supply plans, it was decided, ought to have the following characteristics:

1) The authorities specify net output levels of final goods (including capital goods and planned inventory accumulations).

2) The system yields a plan which determines, at the very least, a set of output targets for each producing unit.

3) The system yields a plan which is internally consistent. For each output target specified by the plan, sufficient quantities of the necessary inputs must be provided.[5]

4) The system yields a plan which is reasonably efficient in the short run. To the extent permitted by long-term constraints, existing resources and factors of production should be put to full and reasonable use. If the necessary labor is available, a short-term plan ought to set output targets at levels which utilize the full capacities of most producing units. The Morozhenoens did realize, however, that the scope of an annual plan is too limited to deal with questions of efficiency involving the allocation of capital or major shifts in technology.

II. The Input-Output Planning Procedure

The Morozhenoens have had a long standing aversion to Russian bureaucratic planning methods. In 1947, risking Moscow's ire, they invited a well-known American economist from a major New England academic institution to design for them a simple, non-bureaucratic material supply planning system. They stipulated that the system have the previously discussed characteristics. The American advised the use of the following input-output procedure.

Let y_j be the gross output target for commodity j, d_j the final demand for commodity j, and a_{ij} the input of commodity i required per unit output of commodity j (a fixed-proportions production function is assumed), and let Y, D, and A denote the corresponding vectors and matrix, respectively. Suppose that A and D are known. To determine a consistent set of gross output targets, gross supplies are equated with gross demand:[6]

$$Y = AY + D \tag{1}$$

Solving for Y, we obtain

$$Y = (I - A)^{-1}D \tag{2}$$

where I is the identity matrix. The vector Y may also be given inductively, as in the following:

$$Y_{(i)} = AY_{(i-1)} + D \tag{3}$$

where $Y_{(i)}$ approaches Y as i gets large.

Equation (3) specifies what we shall call a "supply-demand iteration"; it sets a new output (supply) vector equal to the demand vector associated with a previous output vector. To use (3), one starts out by specifying a tentative gross output vector, say $Y_{(0)}$. The right-hand side of (3) yields the demand vector associated with the production of $Y_{(0)}$; namely, $AY_{(0)}+D$. A new tentative output vector $Y_{(1)}$ would then be defined as equal to this demand vector. Equation (3) would be applied repeatedly, yielding $Y_{(2)}$, $Y_{(3)}$, and so on.

---

[5] This is only one sense in which a plan may be said to be consistent. Richard Stone provides a fairly complete summary of the concept of consistency in planning.

[6] In this paper, the terms "output" and "supply" are used interchangeably. Imports and planned decreases in inventory stocks are considered to be negative summands of final demand. The demand for an intermediate good is completely determined by the output targets and the fixed-proportions production functions and is assumed to be strictly independent of price.

If any given output vector $Y_{(i)}$ were produced, then the corresponding demand vector would be $AY_{(i)}+D$. Let $E_{(i)}$ be defined as the supply-demand imbalance associated with the production of $Y_{(i)}$, that is, the difference between the supply $Y_{(i)}$, and the demand $AY_{(i)}+D$. Then, the imbalance associated with $Y_{(i)}$ is given by

$$E_{(i)} = A^iE_{(0)}$$

Since A is productive,[7] we know that A^i approaches 0 as the exponent i gets large. It follows that with succeeding supply-demand iterations $E_{(i)}$ also approaches 0 and that $Y_{(i)}$ goes to Y. (How fast does the supply-demand imbalance approach 0? A trial with a 38-sector A-matrix for the USSR and with $E_{(0)}$ proportional to gross output, revealed that in general, each iteration cut the supply-demand imbalance roughly in half.)[8]

If the demand for final products, D, and the input norms, A, are known to a central planning agency, the agency could grind out Y with a computer in a relatively short period of time using either (2) or (3). It would remain to disaggregate the gross output targets with respect to individual enterprises. In Morozhenoe, however, because of the relatively small number of enterprises, this would present few difficulties.

The input-output procedure is extremely attractive because of its simplicity and because plans produced by it are perfectly consistent. And while the above planning method does not imply short-run efficiency (which depends on the values of A and D), it is not incompatible with it. Unfortunately, the input-output procedure could not be successfully implemented in Morozhenoe.

The principal difficulty with the scheme was that all information on A and D had to be known by the center in order to be used. The relevant information had to be obtained in great detail at the local level and then summed up for the entire economy, a task which overwhelmed Morozhenoen information processing facilities. Even for this small 1000-good economy, there are a total of 1,001,000 entries in A and D, of which roughly 200,000 are non-zero.[9] And since the amount of data to be compiled grows more than proportionally to the number of commodities produced, the difficulty of this task increased rapidly as the economy grew and became more complex.[10]

---

[7] A technology matrix A is said to be productive if a positive vector of net outputs is producible; i.e., A is productive if, and only if, there exists Y such that $Y-AY>0$ (in every component). It can be shown that A is productive if, and only if, $(I-A)^{-1}$ exists and is nonnegative.

[8] A 38-sector A-matrix and a vector Y of gross outputs for the Soviet Union were obtained from V. Treml (1966). Starting with supply-demand imbalances proportional sector-by-sector to gross outputs, successive supply-demand iterations (3)—or, equivalently, successive multiplications of the imbalance vector by the A-matrix—reduced the ruble value of the supply-demand imbalances by an average of 51.0 percent per iteration. In the sector with the smallest average reduction, the imbalance was reduced by 28.4 percent per iteration. Had we assumed that the imbalances in different sectors had differing signs, the reductions would have been considerably greater. (Note: successive iterations do not reduce the imbalance in a given sector by a constant percentage; the above figures are arithmetic averages, with each percentage reduction weighted by the size of the imbalance before the iteration.)

It is unlikely that the degree of aggregation of the A-matrix and the imbalance vector would have a significant effect on the rate at which multiplication of the imbalance vector by the A-matrix reduces that vector. The above experiment was repeated with a 17-sector version of Treml's matrix, and the results were virtually unchanged. Indeed, if we assume that the A-matrix was derived from an input-output table, and if the original imbalance vector is proportional to the gross output vector given in that table, then it can be shown that the percentage change in total value of the imbalances as a result of multiplication by the A-matrix is independent of the degree of aggregation.

[9] A crude estimate. In Eidel'man (p. 237), it is reported that 4,754 non-zero entries were contained in 83-sector Soviet 1959 A-matrix. Assuming that the number of non-zero entries in an A-matrix is greater than proportional to the dimension of the matrix and less than proportional to the number of entries in the

By 1953, trial use of the input-output planning procedure was discontinued. The stock of unwanted inventories equaled almost two-thirds the size of the annual gross national product, and shortages abounded. The leading Morozhenoen economists realized that this situation resulted not from any error internal to the planning procedure, but rather from the use of incorrect values in the A-matrix and in the final demand vector D. Some economists argued that the techniques of information gathering should be improved and that the input-output planning system be retained. But they were overruled by the authorities who insisted that the planning system be scrapped in favor of a new system that requires much less information in order to function. The Morozhenoen government set up an emergency committee to draw up an annual planning procedure based on Russian practices at the time. On October 20, 1953, this new procedure was formally adopted.

III. The Soviet-Type Planning Procedure in Morozhenoe

The Soviet-type planning procedure depends on the existence of a large planning bureaucracy, both at the central and at the enterprise levels. All commodities are placed in either of two categories: centrally planned commodities and locally planned commodities. Information on locally planned commodities is processed entirely at the enterprise level, while information pertaining to centrally planned commodities is passed back and forth between the central planning agency and the enterprises.

In Soviet Morozhenoe, the central planning agency is known as Gosplan. Gosplan has a series of departments organized along industrial lines. Each centrally planned commodity is assigned to two of these departments: a so-called summary department, and an industrial department. The summary departments are responsible for processing information dealing with demand and distribution of commodities, while industrial departments deal with the supply of commodities from production and other sources. Gosplan also contains statistical departments and final-demand departments. How the various departments of Gosplan interact with each other and with local planning agencies is described on the following pages.

Conceptually speaking, the Soviet-type planning procedure may be thought of as a series of iterations. These iterations are of three types, all of them variants of the basic supply-demand iteration described by equation (3). One of these, the "retrospective iteration," may be used by local planning agencies in determining output targets for locally planned commodities, and by central planning agencies in determining output targets for centrally planned commodities. The other two, the "external iterations" and the "internal iterations," require certain centrally processed information, and, consequently, may be used to adjust the output targets of only centrally planned commodities. In Morozhenoe, the process of drawing up an annual plan spans a one-year period.

The principal steps of the annual planning procedure are outlined below (see, also, Figure 1) and a mathematical representation for the procedure is presented.

Let n denote the number of commodities

---

matrix, a guess of 200,000 as the number of non-zero elements in the 1000x1000 Morozhenoen A-matrix was derived by taking a geometric average of 4754x1000/83 and $4754\text{x}(1000/83)^2$.

[10] The difficulties of applying input-output methods to Soviet planning are outlined by A. Dorovskikh (pp. 38–39).

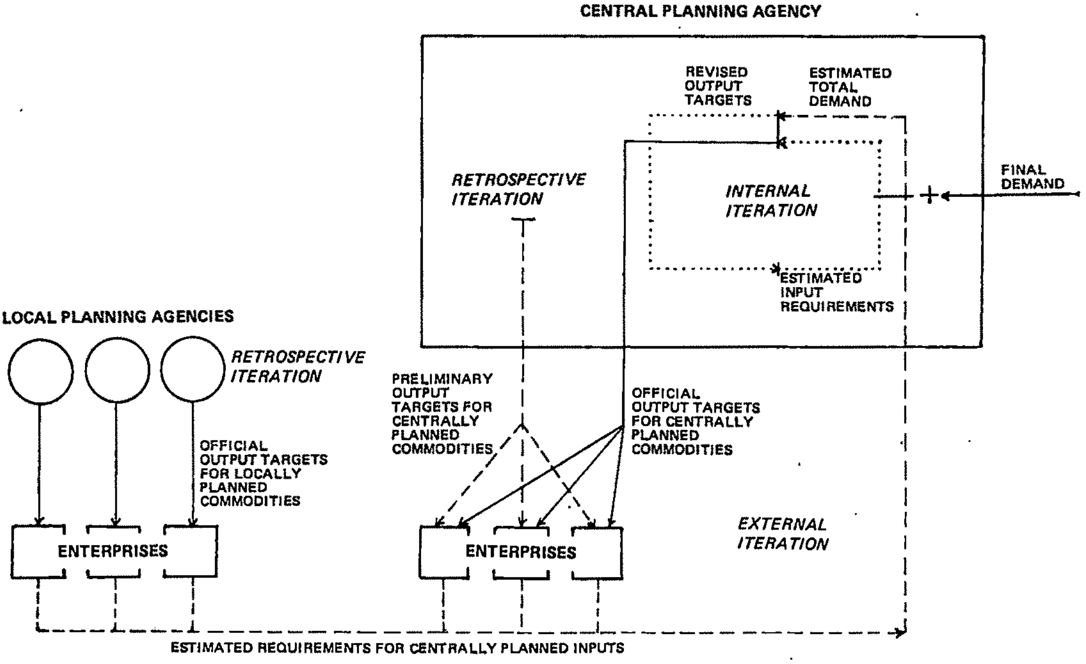

CENTRAL PLANNING AGENCY
REVISED OUTPUT TARGETS
ESTIMATED TOTAL DEMAND
RETROSPECTIVE ITERATION
INTERNAL ITERATION
FINAL DEMAND
ESTIMATED INPUT REQUIREMENTS
LOCAL PLANNING AGENCIES
RETROSPECTIVE ITERATION
PRELIMINARY OUTPUT TARGETS FOR CENTRALLY PLANNED COMMODITIES
OFFICIAL OUTPUT TARGETS FOR CENTRALLY PLANNED COMMODITIES
OFFICIAL OUTPUT TARGETS FOR LOCALLY PLANNED COMMODITIES
ENTERPRISES
ENTERPRISES
EXTERNAL ITERATION
ESTIMATED REQUIREMENTS FOR CENTRALLY PLANNED INPUTS

FIGURE 1—SOVIET-TYPE MATERIAL-SUPPLY PLANNING PROCEDURE: A STYLIZED VERSION

being produced. All of the vectors defined below are n-dimensional; the ith component of a vector always pertains to commodity i. If the vector is subscripted, the subscript refers to the year for which the vector is defined. Each commodity is produced by many enterprises. The word "aggregate" is used to indicate a nationwide total pertaining to a given commodity, and *not* the combining of different commodities. A list of symbols used appears in the Appendix.

Step 1: Specification of preliminary aggregate output targets: a retrospective iteration (January–March).[11]

For each centrally planned commodity, the center sets a preliminary aggregate output target by adding an exogenously determined increment in gross output to the previous year's aggregate demand.[12] The preliminary aggregate output targets (control figures) are then used to assign output targets to individual enterprises.

No preliminary aggregate output targets are set for locally planned commodities. Instead, a preliminary output target is set independently for each enterprise producing a given commodity, by adding an exogenously determined increment in production to the demand for the commodity on the enterprise (i.e., the orders placed with the enterprise) during the previous year. The reader should note that while a preliminary output target is not set, or known, for a locally planned product, such a target does in fact exist—namely, the sum of all of the enterprise output targets for the product.

For both centrally and locally planned

---

[11] The time periods given here are similar to those for the Soviet Union as reported in Evenko, (p. 84).

[12] It is assumed that in cases where actual requirements for centrally planned inputs differ from planned allotments, producers will ask the center to make appropriate adjustments in the allotments. These requests for adjustments enable the center to calculate actual aggregate demand for centrally planned products after the fact.

commodities, preliminary output targets (potential supplies) are based on actual demands of the previous year. Thus the designation retrospective iteration.

Let $\hat{Y}_t$ be the vector of preliminary aggregate output targets for year t. Then, in accordance with the above specifications, $\hat{Y}_t$ is given by the vector equation

$$\hat{Y}_t = X_{t-1} + Q_t$$

where Q_t is a vector of exogenously determined increments (or decrements, if negative) to aggregate outputs,[13] and X_t is the vector of aggregate demands. For analytical purposes, however, we shall allow the possibility of omitting the retrospective iteration for some commodities, in which case the previous year's output would be substituted in the above equation for the previous year's demand. Therefore we amend the equation to

$$\hat{Y}_t = \Omega X_{t-1} + (I - \Omega) Y_{t-1} + Q_t \tag{4}$$

where Y_t is the vector of aggregate outputs, Ω is a diagonal matrix whose ith entry is equal to 1 if commodity i is to be iterated retrospectively, and 0 otherwise. All quantities indicated in (4) are aggregate quantities.

Equation (4) is valid for both centrally and locally planned commodities, but for a different reason in each of the two cases. It is true for centrally planned commodities because the right-hand side of the equation actually describes the way the center sets preliminary aggregate output targets. Equation (4) is valid for locally planned commodities, because an equation analogous to (4) but in terms of nonaggregate quantities, defines each enterprise target. Summing all of the nonaggregate equations pertaining to a particular locally planned commodity yields the relationship for aggregate quantities expressed by (4).

Step 2: Computation of tentative demands (April–June). The vector $\hat{X}_t$ of tentative aggregate demands induced by the preliminary output targets $\hat{Y}_t$, is given by

$$\hat{X}_t = A\hat{Y}_t + D_t \tag{5}$$

where D_t is the vector of planned final demands. Gosplan, however, cannot use this equation to compute the tentative demands, since it does not necessarily know the value of the A-matrix or the values of $\hat{Y}_t$ and D_t pertaining to locally planned commodities. Instead, Gosplan calculates aggregate demands for centrally planned products only. The following procedure is used:

All enterprises submit estimates of the quantities of all *centrally planned* commodities they plan to use as inputs.[14] In making these estimates, each enterprise assumes its output will equal its preliminary output target. The estimates of required inputs are aggregated and forwarded to Gosplan. At the same time, final demands for centrally planned products are set by final demand departments of Gosplan, which, at its discretion, may be guided by consumers' and investors' desires.[15] Each summary department of Gosplan then determines estimates of aggregate demands for the centrally planned products assigned to it by adding the demand for each product as an input to the respective final demand.

Step 3: Revision of the preliminary output targets for centrally planned commodities: an external iteration (July).

At this point, the summary departments of Gosplan consult with the corresponding industrial departments. If the tentative demands do not equal the tentative supplies—and they surely will not—then the demand and supply figures must be reconciled. In order to effect this reconciliation, Gosplan uses the material balance, an accounting device in which all supplies of a commodity, on the one hand, and all demands for the commodity, on the other, are tabulated. In most Socialist countries, the process of adjusting supplies and demands in order to bring about a balance is extremely complicated and draws heavily on the experience and intuition of the planners. In Soviet Morozhenoe, however, only one basic aspect of the process is used. The preliminary output targets for centrally planned commodities are reset to equal the respective aggregate demand estimates. Thus a new tentative aggregate gross output target is formulated for each centrally planned commodity.

---

[13] The centrally planned components of Q_t are set explicitly by Gosplan; the locally planned components are the sums of exogenous changes made by local planning agencies.

[14] For a description of this process in the USSR, see Levine (1959, pp. 156–59).

[15] The planning of consumption is treated at length by Phillip Weitzman.

The output targets for locally planned commodities remain unchanged.

Let $Y_{t(0)}$ be the revised vector of aggregate gross output targets. It follows from the above that $Y_{t(0)}$ will equal X_t in centrally planned components and $\hat{Y}_t$ in locally planned components. These relationships can be expressed as one equation:

$$Y_{t(0)} = \Gamma \hat{X}_t + (I - \Gamma)\hat{Y}_t \tag{6}$$

where Γ is a diagonal matrix with entries pertaining to centrally planned commodities set equal to 1, and entries pertaining to locally planned commodities set equal to 0.

As far as centrally planned commodities are concerned, Steps 2 and 3 have the same result as one supply-demand iteration (3) of the input-output procedure, in that the demands generated by the preliminary output targets become new output targets. With the input-output procedure, however, the center must calculate the demands for inputs from its knowledge of production functions (the A-matrix), while in Step 2 of the Soviet-type procedure, the center (Gosplan) allows the producing units (enterprises) to estimate their own input demands. Because part of the demands are estimated outside of the center, Step 3 is referred to as an external iteration.

Step 4: Further revisions of tentative output targets for centrally planned commodities:[16] internal iterations (August).

In principle, the internal iteration is a straightforward process. Whenever the tentative aggregate output targets for centrally planned commodities are revised, the tentative input requirements of the producers of centrally planned commodities will change. When an industrial department of Gosplan changes (or is advised of a change in) output targets of commodities under its purview, it may advise appropriate summary departments of resulting changes in the input requirements of the industry, provided, of course, that the information necessary to compute the changes in requirements is available. After the summary departments receive notification of changes in input requirements from all of the industrial departments, the summary departments may proceed to calculate a revised set of aggregate demands for the centrally planned commodities. The summary departments may then consult with the corresponding industrial departments, so that the latter can set new aggregate output targets equal to these revised aggregate demands.

Let $Y_{t(i)}$ denote the tentative aggregate output targets after i internal iterations. Then we may represent an internal iteration by

$$Y_{t(i)} = \Gamma X_{t(i-1)} + (I - \Gamma)\hat{Y}_t \tag{7}$$

where $X_{t(i)}$, the vector of tentative aggregate demands after *i* internal iterations, is given by

---

[16] That a realistic model of Soviet-type planning ought to include this step was suggested by Martin Weitzman, Yale University, and by Maria Augustinovics, Head, Division of Long-term Planning, National Planning Institute, Budapest, Hungary.

$$X_{t(i)} = AY_{t(i)} + D_t \tag{8}$$

The internal iteration also yields results like those of the supply-demand iteration (3), except that in this case, only the output targets of centrally planned commodities are revised. The name internal iteration derives from the fact that all computation is performed within Gosplan, itself. In order for the internal iteration to be carried out, the industrial departments must have certain information about input usage in the production of commodities under their purview. In particular, they must know the specific coefficients in the A-matrix which yield the quantities of centrally planned inputs needed to produce these commodities.

Step 4 consists of a number of internal iterations, a number bounded by time, cost and information constraints on Gosplan. Because of these constraints, some of these internal iterations may have to be partial,[17] or all internal iterations may even be omitted.

At this point, the reader may be wondering why an external iteration is necessary given that subsequent iterations may be internal. The explanation lies in the fact that the first calculation of tentative aggregate demands for centrally planned commodities requires more information than subsequent calculations do. The first calculation requires information on centrally planned inputs used to produce *all* commodities, both centrally and locally planned. But, because the tentative output targets for centrally planned commodities are the only ones that change in the iterations, subsequent calculations of the tentative aggregate demands require information pertaining only to the use of centrally planned inputs in the production of centrally planned commodities. Furthermore, the center may be able to obtain much of the information needed to execute internal iterations as a by-product of the external iteration.

Step 5: Specification of the official output targets (September–December).

The most recently derived tentative aggregate gross output targets for centrally planned commodities become official targets after approval by various government bodies. These targets are then disaggregated and assigned to the individual enterprises producing centrally planned commodities. Official output targets for locally planned goods are the same as the preliminary targets set in Step 1.

If *n* internal iterations were carried out, the official output targets $\overline{Y}_t$ are given by

$$\overline{Y}_t = Y_{t(n)} \tag{9}$$

It is clear that the Soviet-type planning procedure is less demanding with regard to information gathering than is the input-output procedure. The input-output procedure required the center to know all of the elements of the A-matrix, to calculate total output and demand for each commodity, and to know or set total final demand for each commodity. The Soviet-type procedure, however, requires only that the center calculates total output and demand for centrally planned commodities and that it knows or sets final demand for centrally planned commodities. If internal iterations are omitted, knowledge of the A-matrix by the center is not required. However, if the center does know coefficients pertaining to the use of centrally planned inputs in the production of centrally planned commodities, it may use them to carry out internal iterations, thereby increasing the degree of consistency of the plan—as we shall see.

---

[17] For the sake of mathematical simplicity in analyzing the model, I have assumed that an internal iteration involves all centrally planned commodities. A more general model would allow internal iterations to be performed for any subset of the centrally planned commodities. This would permit the center to carry out partial internal iterations which would require the knowledge of technical coefficients only for those commodities involved.

As the Soviet-type procedure was being established in Morozhenoe, many economists there predicted that plans produced by the procedure would be subject to catastrophic failure. It was agreed that the system does allow (but does not require) the authorities to set the level of final demand for a commodity, and does yield output targets for each producing unit. There was some concern about short-run economic efficiency. But the major cause of skepticism centered around the question of consistency.[18] Why, it was asked, should the planned supply of a commodity equal the resulting demand for it? For one thing, only about one-third of all commodities were to be centrally planned. Considering the fact that current output targets for locally planned goods are not directly coordinated with the current output targets for centrally planned goods, what assurances are there that the supplies of locally planned goods required as inputs throughout the economy would be available? The usefulness of the external and internal iterations seemed doubtful for two reasons: the planners would have time and funds available for at most a few iterations, and only centrally planned output targets were being revised. It was suggested that the abolition of all central planning would be preferable to these halfway measures.

Between 1954 and 1956, the chaotic state of the Morozhenoen economy seemed to bear out the pessimism of the economists. But, surprisingly, the plan of 1957 turned out to be almost consistent. The supply and demand for each centrally planned product differed at most by a few percentage points, and the supply and demand for each locally planned product differed by no more than about 5 percent. And this, despite the fact that internal iterations had been entirely omitted from the planning procedure for economy reasons. Results in later years were just as good. An analysis of the model suggests an explanation.

IV. An Analysis of the Soviet-Type Procedure

It is assumed that the relationship between actual outputs Y_t, demands for the outputs X_t, and final demands D_t, is given by

$$X_t = AY_t + D_t \tag{10}$$

The vector of supply-demand imbalances E_t is given by

$$E_t = Y_t - X_t \tag{11}$$

For the purposes of this analysis, we make the simplifying assumption that the planned output targets are always exactly fulfilled, i.e.,

$$Y_t = \overline{Y}_t \tag{12}$$

This implies that E_t will be generated only by the inconsistencies in the plan.

The assumption that output targets are precisely met also implies that the supply-demand imbalances represented by E_t manifest themselves as economic occurrences with little or no secondary effects: perhaps as unplanned changes in the levels of consumption,[19] in the levels of exports or imports, or, under certain circumstances, in the size of inventory stocks. In other words, the entire error falls on final demand. Although these assumptions are unrealistic (even in Morozhenoe), they serve to isolate the direct effects of inconsistent plans from the secondary effects, which depend to a large extent on how the economic administration handles the shortages or surpluses that arise from the inconsistencies.

---

[18] Problems of feasibility are avoided by assuming that the Morozhenoens always choose final demands so as to fall within the production possibility curve. Thus if a plan is internally consistent, it is also feasible. The problem of feasibility is treated more realistically in the author's doctoral dissertation.

[19] Of course, unplanned (or any other) changes in the level of consumption will have important secondary effects outside the sphere of production. And if the changes are large enough, or if the population is close to the subsistence level, changes in consumption may effect labor, and consequently, production.

The vector E_t tells the complete story as far as consistency of the plan is concerned. If $E_t=0$, then the plan is perfectly consistent, if the components of E_t are large (+ or −), then large surpluses and/or shortages are indicated. We attempt to throw some light on the genesis of E_t in the Morozhenoen planning procedure by solving the mathematical representation of that procedure. For simplicity, we shall assume that no internal iterations (Step 4) are carried out. Under this circumstance, equations (6)–(9) and (12) reduce to

$$Y_t = \Gamma \hat{X}_t + (I - \Gamma)\hat{Y}_t \tag{13}$$

The model of the planning procedure with internal iterations is analyzed in the Appendix.

Equations (4), (5), (10), (11), and (13) can be solved simultaneously for the supply-demand imbalance E_T realized in a given year T. If, as we may do without loss of generality, we set $Y_0=D_0=0$, the solution is as follows:

$$E_T = \sum_{t=1}^{T} (\Gamma_{[1]}\Omega_{[1]})^{T-t}\Gamma_{[1]}L_t \tag{14}$$

where

$$L_t = Q_t - (AQ_t + \Delta D_t) \tag{15}$$

$$\Gamma_{[1]} = I - (I - A)\Gamma \tag{16}$$

and

$$\Omega_{[1]} = I - (I - A)\Omega \tag{17}$$

and where $\Delta D_t = D_t - D_{t-1}$.

Equation (14) has a helpful heuristic interpretation. Note that (14) gives the supply-demand imbalance E_T for year T as a weighted sum of L_t's. What does L_t represent? Recalling that Q_t is the exogenously determined increment in gross output for year t (see equation (4)), we see that $AQ_t+\Delta D_t$ is the exogenously induced increment in aggregate demand. It follows from (9), then, that L_t is the magnitude of the supply-demand imbalance newly generated in year t by disproportions between the exogenously induced increment in gross output, on the one hand, and the exogenously induced change in demand, on the other. Thus, equation (14) breaks down the supply-demand imbalance E_T into a sum, and each term $(\Gamma_{[1]}\Omega_{[1]})^{T-t}\Gamma_{[1]}L_t$ of the sum may be thought of as that portion of the imbalance that was generated in year t.

It turns out that any reasonable specification of the parameters Γ and Ω will guarantee that $(\Gamma_{[1]}\Omega_{[1]})^{T-t}\Gamma_{[1]}$ approaches 0 geometrically, as T−t increases. (Setting $\Omega=I$, its normal value, is sufficient to assure this.) It follows that changes made long ago do not contribute significantly to the current imbalance. Given a uniform degree of consistency of the exogenous increments to output and demand, the degree of consistency of the annual plans will remain at a kind of stable equilibrium. There is no positive feedback endogenous to the system which could result in a deterioration of the consistency of the plans from year to year. It is possible, of course, that the consistency of the exogenous increments to output and demand could steadily worsen over time, i.e., the differences between the increments to output and increments to demand could become a larger and larger percentage of the size of the increments to output. But even relatively large imbalances in the exogenous increments to output and demand, do not generate major inconsistencies in the annual plans (see examples in Section VI).

The size of the imbalance in the exogenous increments to output and demand is not the only determinant of the degree of consistency in the plans. The degree of consistency also depends on how many and which goods are centrally planned and on the number of internal iterations in the planning procedure, as we show below.

V. Some Specific Solutions for Plan Inconsistencies

What does (14) look like for various values of the parameters Ω and Γ? If a retrospective iteration is performed for all commodities ($\Omega=I$), then with no central planning at all ($\Gamma=0$), we have

$$E_T = L_T + AL_{T-1} + A^2L_{T-2} + \dots + A^{T-1}L_1 \tag{18}$$

In this case, the portion of the supply-demand imbalance of year T that was generated in year t is $A^{T-t}L_t$. Thus, since A is productive, the longer the elapsed time from the introduction of exogenous increments in gross output and final demand, the smaller, in general, will be the supply-demand imbalances attributable to those changes. (For an actual A-matrix, multiplication of a vector by A^i reduced most components of that vector by a factor of about 2^i (see footnote 8). E.g., the part of a supply-demand imbalance generated by output and demand increments made five years previously would be less than roughly 3 percent of the imbalance that would have been caused by the same increments at the time they were introduced.)

This solution bears an interesting relationship to the supply-demand iteration procedure (3). Suppose that supply-demand iterations were used to balance the exogenously determined increment in supply Q_t with an exogenous increment in final demand ΔD_t. From (3) we get

$$Q_{t,i} = AQ_{t,i-1} + \Delta D_t$$

where $Q_{t,i}$ is defined to be Q_t as revised by i iterations (with $Q_{t,0}=Q_t$). Solving this as a difference equation yields

$$Q_{t,i}=A^iQ_t+(I+A+\dots+A^{i-1})\Delta D_t \tag{19}$$

Now, if, for a given i, the output increment $Q_{t,i}$ alone were produced, the supply-demand imbalance resulting from that production would be given by

$$Q_{t,i} - (AQ_{t,i} + \Delta D_t)$$

which, by (19), equals

$$A^i[Q_t - (AQ_t + \Delta D_t)]$$

and, by (15), this reduces to A^iL_t. But this term has the same structure as the terms on the right-hand side of (18). Thus, the Soviet-type planning procedure under these assumptions works as if it treats the supply and demand of a given year, not as a whole but as a series of annual exogenous increments which have accumulated over time. And it is as if the planning procedure uses supply-demand iterations to balance each set of annual increments separately from the other sets, with the number of iterations carried out in each case equal to the number of years having passed since that set of increments were introduced.

Suppose, now, that all commodities are centrally planned ($\Gamma=I$), but that the retrospective iteration is completely omitted ($\Omega=0$). Then

$$E_T = AL_T + A^2L_{T-1} + \dots + A^TL_1 \tag{20}$$

This procedure works much the same way as the procedure associated with (18), except that the balancing process for each set of increments includes one more supply-demand iteration than is the case in (18). And this is not surprising. Under the assumptions pertaining to (18), the whole planning procedure amounts to a single retrospective iteration for all commodities. Under the assumption of central planning for all commodities, which yields (20), the procedure amounts to a single external iteration for all commodities. And the only difference between the retrospective iteration and the external iteration is that the latter takes account of the *currently planned* increments in output and final demand while the former does not.

When the planning procedure specifies central planning for all commodities (i.e., an external iteration) on top of a complete retrospective iteration ($\Gamma = I$, $\Omega = I$), the resulting supply-demand imbalance for year T will be given by

$$E_T = AL_T + A^3L_{T-1} + A^5L_{T-2} + \dots + A^{(2T-1)}L_1 \tag{21}$$

It is as if each annual set of exogenous increments were balanced with two supply-demand iterations per year, instead of one. The incorporation of central planning for all commodities into the planning procedure more than halves the number of years it takes a newly generated supply-demand imbalance L_t to damp out.

When the planning procedure specifies central planning for all commodities and includes internal iterations, it is as if each annual set of exogenous increments were subject to even more supply-demand iterations per year (see the Appendix). If the planning procedure included three internal iterations, for example, the supply-demand imbalance E_T would be given by

$$E_T = A^4L_T + A^9L_{T-1} + A^{14}L_{T-2} + \dots + A^{(5T-1)}L_1 \tag{22}$$

While the above cases illustrate the properties of the retrospective and the external and internal iterations, a more realistic example of the Soviet-type procedure is provided by assuming central planning only for some commodities on top of a complete retrospective iteration in setting the initial tentative targets. In this case, $\Omega = I$, but Γ would contain both 1's and 0's on its diagonal. For simplicity we again assume that no internal iterations are performed. With these assumptions, (14) yields the following solution for E_T:

$$E_T = \tilde{A}L_T + (A\tilde{A})\tilde{A}L_{T-1} + (\tilde{A}A)^2\tilde{A}L_{T-2} + \dots + (\tilde{A}A)^{(T-1)}\tilde{A}L_1 \tag{23}$$

where $\tilde{A}$ is constructed by taking columns corresponding to centrally planned commodities from the A-matrix, and columns corresponding to locally planned commodities from the identity matrix.

Equation (23) implies the obvious fact that each annual set of increments in output and demand for centrally planned commodities is subject to one more supply-demand iteration than those of locally planned commodities, and it follows that even without using internal iterations, the supply-demand imbalance of a centrally planned commodity would be considerably smaller than that of the commodity were it locally planned. Equation (23) also implies the less obvious fact that the central planning of some commodities generally reduces the supply-demand imbalances of the other commodities which remain locally planned.

VI. Some Illustrative Numerical Solutions for Plan Inconsistencies

Having explored the properties of solutions for the supply-demand imbalance under several specifications of the planning procedure, we try to get more of a hold on the size of the indicated supply-demand imbalances by using real data and a few assumed numbers. In what follows, we assumed that the annual exogenous increments in gross output Q_t are on the order of 10 percent of the total gross output, and that the supply-demand imbalances inherent in these changes L_t are about 25 percent of the size of the increments, and we used Treml's 38-sector version of the 1959 Soviet A-matrix to evaluate our formulas for E_T. In evaluating these formulas, T was chosen sufficiently large to make E_T converge to its steady-state value.

For the case where the planning procedure incorporates only a retrospective iteration, we learned, by evaluating (18), that the supply-demand imbalances average 4.9 percent of the size of the corresponding gross outputs. The sector with the largest imbalance had an imbalance of 8.8 percent.

When all commodities are retrospectively iterated and centrally planned (with no internal iterations), the imbalances average about 1.6 percent of the size of the gross outputs (from (21)); and the addition to the planning procedure of three internal iterations lowers the size of the imbalances to an average of .14 percent of the gross outputs (from (22)).

With half of the sectors centrally planned (no internal iterations used), the supply-demand imbalances of the centrally planned sectors average 1.7 percent of the size of the respective gross outputs as compared to an average of 4.5 percent for the same sectors with no central planning. Also, centrally planning half of the sectors lowers the imbalances in the remaining sectors (still locally planned) from an average of 5.4 percent to an average of 4.9 percent.[20]

VII. Conclusion

In order to gain some insight into the workings of the Soviet-type system of material supply planning, I have constructed a model of "Morozhenoen" planning, which is, of course, a stylized abstraction of the Soviet system, itself. The model was designed with several well-known descriptions of the Soviet procedure in mind, including those of Levine (1959, 1961), Montias (1959), and Evenko. The input-output procedure, which was explicated for the purpose of comparison with the Soviet-type system, is standard.

The input-output method of setting output targets is very attractive because in either its matrix-inversion form (2), or its iterative form (3), virtually perfect balance between supply and demand can be achieved, at least in theory. Unfortunately, in order to use the input-output procedure for detailed material supply planning, the center must have available to it a large amount of information which is difficult to obtain and process. This fact renders the input-output planning procedure of doubtful value to economies which lack advanced facilities for data collection and processing.

The traditional Soviet-type material supply planning procedure, on the other hand, takes advantage of detailed information known on the local level but not known by the central planning agency. Information concerning only "locally planned" commodities need not be known to anyone outside of the producing enterprise. And even with regard to centrally planned commodities, much less information is needed by the center with the Soviet-type procedure than would be needed with an input-output procedure. Yet, as I hope the model has shown, the Soviet-type procedure can, in principle, yield a reasonably consistent plan of material supply. Moreover, as Montias (1959, pp. 967–68) argued years ago, the underlying workings of the system are similar to the iterative version of the input-output procedure defined by (3).

The abstract Soviet procedure as I defined it, was shown to be operationally equivalent to a succession of three kinds of iterations. In the first of these, the retrospective iteration, preliminary output targets are set on the basis of actual demands of the preceding year. This iteration may be carried out entirely on the local level for locally planned commodities and entirely on the central level for centrally planned commodities. In the external iteration which follows, revised output targets for centrally planned commodities are set on the basis of demand estimates made in part on the local level assuming the original preliminary output targets. Finally, there may be some internal iterations in which revised output targets for centrally planned commodities are calculated within the planning center itself.

---

[20] In this example, the largest percentage imbalance among the nineteen centrally planned sectors is 5.7 percent, and the largest among the nineteen locally planned sectors is 6.9 percent.

But how can only a few iterations yield a reasonably consistent plan? The answer is that the effect of the iterations accumulates from year to year, a phenomenon brought about by the fact that each year's planning procedure uses previous plans and production and demand figures as a point of departure. As is reflected in the solutions for the supply-demand imbalance derived in the previous section, it is only the recent changes in output and demand that are iterated only a few times. The bulk of these magnitudes were iterated over and over again in previous years. This point, I think, was missed by Levine when he stated the following:

> It is frequently thought that the iterative approach is the basic method used by Soviet planners to achieve consistency in their plans. I do not think that this theory is correct. . . . On the basis of *very crude* calculations it might be said that somewhere between 6 and 13 iterations would be required. It is inconceivable that Gosplan, under the conditions which prevailed could have performed that number of iterations.[21]

Of course, as Levine asserts elsewhere in his article, Gosplan may well rely strongly on techniques of balancing which are not akin to supply-demand iteration, the tightening of input coefficients and the substitution of available inputs for scarce ones, for example. But I have established with the model that, in theory, at least, a basically iterative procedure of material supply planning can work.

For the record, let me note that this essay omits a number of important questions. For one thing, the model developed here does not provide for many institutional factors in Soviet-type economies which typically contribute to the inconsistencies of a plan. In addition, how capacity limitations enter into the planning procedure was not discussed. Nor was the problem of holding inventories at a reasonable level explored. These last two matters are treated in Manove (pp. 110–258), where a simulation of a version of this model and of an algorithm for nonprice rationing of intermediate goods is described.

APPENDIX

Solution of the Model of the Soviet-Type Planning Procedure with Internal Iterations

List of Symbols:

A = technology matrix—coefficients of input per unit output (matrix)
Q_t = exogenously determined increments to aggregate outputs in year t (vector)
D_t = final demands in year t (vector)
Y_t = gross outputs in year t (vector)
X_t = aggregate demands in year t (vector)
E_t = supply-demand imbalances in year t (vector)
$\hat{Y}_t$ = preliminary aggregate output targets for year t (vector)
$\hat{X}_t$ = initial tentative aggregate demands for year t (vector)
$Y_{t(i)}$ = tentative aggregate output targets after i internal iterations (vector)
$X_{t(i)}$ = tentative aggregate demands after i internal iterations (vector)
$\bar{Y}_t$ = official aggregate output targets for year t (vector)
Ω = retrospective iteration parameter: elements pertaining to commodities retrospectively iterated are set to 1,

---

[21] Levine (1959, p. 165). Levine reasserts this position in a later paper (1966, p. 273).

other elements set to 0 (diagonal matrix)

Γ = central planning parameter: elements pertaining to centrally planned commodities set to 1, other elements set to 0 (diagonal matrix)

I = identity matrix

N = number of *ex ante* iterations in the planning procedure—i.e., 1 external iteration and N-1 internal iterations (scalar)

T = the current year (scalar)

For $\Gamma^{\{n\}}$, $\Gamma^{[n]}$, $\Gamma_{\{n\}}$, $\Gamma_{[n]}$, $\Omega_{\{1\}}$, $\Omega_{[1]}$, see equations (A.9)–(A.12), (A-18), and (A.19).

Model

Tautological relationships:

$$X_t = AY_t + D_t \tag{A. 1}$$

$$E_t = Y_t - X_t \tag{A. 2}$$

Initial tentative output targets are set (a retrospective iteration):

$$\hat{Y}_t = \Omega X_{t-1} + (I - \Omega)Y_{t-1} + Q_t \tag{A. 3}$$

Aggregate demands are estimated:

$$\hat{X}_t = A\hat{Y}_t + D_t \tag{A. 4}$$

An external iteration (for centrally planned commodities only) is carried out by the center:

$$Y_{t(0)} = \Gamma\hat{X}_t + (I - \Gamma)\hat{Y}_t \tag{A. 5}$$

Internal iterations (for centrally planned commodities only) are carried out by the center. The ith internal iteration yields:

$$Y_{t(i)} = \Gamma X_{t(i-1)} + (I - \Gamma)\hat{Y}_t \tag{A. 6}$$

where

$$X_{t(i)} = AY_{t(i)} + D_t \tag{A. 7}$$

The vector of official output targets $\overline{Y}_t$ is given by $\overline{Y}_t = Y_{t(N-1)}$ where $N-1$ is the number of internal iterations that were carried out. And since it is assumed that official targets are precisely fulfilled, we have

$$Y_t = Y_{t(N-1)} \tag{A. 8}$$

Solution for the model for E_T

We define the following additional symbols:

$$\Gamma^{\{n\}} \equiv \sum_{i=0}^{n} (\Gamma A)^i \tag{A. 9}$$

$$\Gamma^{[n]} \equiv \sum_{i=0}^{n} (A\Gamma)^i \tag{A. 10}$$

$$\Gamma_{\{n\}} \equiv \Gamma^{\{n\}} - \Gamma^{\{n-1\}}\Gamma \tag{A. 11}$$

$$\Gamma_{[n]} \equiv \Gamma^{[n]} - \Gamma\Gamma^{[n-1]} \tag{A. 12}$$

It follows that:

$$A\Gamma^{\{n\}} = \Gamma^{[n]}A \tag{A. 13}$$

$$\Gamma\Gamma^{[n]} = \Gamma^{\{n\}}\Gamma \tag{A. 14}$$

$$\Gamma^{\{n\}} = I + \Gamma\Gamma^{[n-1]}A \tag{A. 15}$$

$$\Gamma^{[n]} = I + A\Gamma^{\{n-1\}}\Gamma \tag{A. 16}$$

These equivalences yield:

$$(I - A)\Gamma_{\{n\}} = \Gamma_{[n]}(I - A) \tag{A. 17}$$

We also define:

$$\Omega_{\{1\}} \equiv I - \Omega(I - A) \tag{A. 18}$$

$$\Omega_{[1]} \equiv I - (I - A)\Omega \tag{A. 19}$$

so that

$$(I - A)\Omega_{\{1\}} = \Omega_{[1]}(I - A) \tag{A. 20}$$

We can now proceed to solve equations (A.1)–(A.8) for E_T. Equations (A.4)–(A.8) yield

$$Y_t = Y_{t(N-1)} = (\Gamma A)^N\hat{Y}_t + \left(\sum_{i=1}^{N} (\Gamma A)^{N-i}\right)(\Gamma D_t + (I - \Gamma)\hat{Y}_t)$$

Applying (A.9) and (A.11) we have

$$Y_t = \Gamma_{\{N\}}\hat{Y}_t + \Gamma^{\{N-1\}}\Gamma D_t \tag{A. 21}$$

Substitution of (A.1) and (A.3) into (A.21) gives us

$$\begin{aligned} Y_t = {} & \Gamma_{\{N\}}\Omega_{\{1\}}Y_{t-1} + \Gamma_{\{N\}}\Omega D_{t-1} \\ & + \Gamma_{\{N\}}Q_t + \Gamma^{\{N-1\}}\Gamma D_t \end{aligned} \tag{A. 22}$$

We now solve for $E_t - \Gamma_{[N]}\Omega_{[1]}E_{t-1}$, which turns out to be a function only of exogenous variables and the parameters. By (A.1) and (A.2), we have

$$E_t = (I - A)Y_t - D_t \tag{A.23}$$

Multiplying (A.23) lagged one period by $\Gamma_{[N]}\Omega_{[1]}$, and subtracting the result from (A.23) yields:

$$\begin{aligned} E_t - \Gamma_{[N]}\Omega_{[1]}E_{t-1} &= (I - A)Y_t - D_t \\ &\quad - \Gamma_{[N]}\Omega_{[1]}(I - A)Y_{t-1} + \Gamma_{[N]}\Omega_{[1]}D_{t-1} \end{aligned}$$

and by using (A.17) and (A.20) we get

$$\begin{aligned} &E_t - \Gamma_{[N]}\Omega_{[1]}E_{t-1} \\ &= (I - A)(Y_t - \Gamma_{[N]}\Omega_{[1]}Y_{t-1}) \\ &\quad - D_t + \Gamma_{[N]}\Omega_{[1]}D_{t-1} \end{aligned} \tag{A.24}$$

But from (A.22) we have

$$\begin{aligned} &Y_t - \Gamma_{[N]}\Omega_{[1]}Y_{t-1} \\ &\quad = \Gamma_{[N]}\Omega D_{t-1} + \Gamma_{[N]}Q_t + \Gamma^{[N-1]}\Gamma D_t \end{aligned}$$

and substituting this into (A.24) yields, after manipulations with (A.9)–(A.20), the following:

$$\begin{aligned} &E_t - \Gamma_{[N]}\Omega_{[1]}E_{t-1} \\ &\quad = \Gamma_{[N]}(D_{t-1} - D_t) + \Gamma_{[N]}(I - A)Q_t \end{aligned}$$

Or, defining $\Delta D_t \equiv D_t - D_{t-1}$, we have

$$\begin{aligned} &E_t - \Gamma_{[N]}\Omega_{[1]}E_{t-1} \\ &\quad = \Gamma_{[N]}(Q_t - (AQ_t + \Delta D_t)) \end{aligned} \tag{A.25}$$

Solving (A.25) as a difference equation in t, we get

$$\begin{aligned} E_T = \sum_{t=1}^{T} &(\Gamma_{[N]}\Omega_{[1]})^{T-t}\Gamma_{[N]} \\ &\cdot(Q_t - (AQ_t + \Delta D_t)) \\ &+ (\Gamma_{[N]}\Omega_{[1]})^{T}E_0 \end{aligned} \tag{A.26}$$

Equation (A.26) is a general solution for E_T when the planning procedure includes one external iteration and $N-1$ internal iterations. When the planning procedure includes no internal iterations ($N=1$), and given that $E_0=0$, the solution reduces to equation (14).

REFERENCES

A. Dorovskikh, "Nekotorie voprosi teori i praktiki mezhotraslevogo balansá," (Some Questions on the Theory and Practice of the Interbranch Balance), *Planovoe Khoziaistvo*, Dec. 1967, *44*, 35–44.

M. R. Eidel'man, *Mezhotraslevoi Balans Obshchestvennogo Produkta*, (The Interbranch Balance of the Social Product), Moscow 1966.

M. Ellman, "The Consistency of Soviet Plans," *Scot. J. Polit. Econ.*, Feb. 1969, *16*, 50–74.

I. A. Evenko, *Planning in the USSR*, Moscow 1962.

H. S. Levine, "The Centralized Planning of Supply in Soviet Industry," in Joint Economic Committee, U.S. Congress, *Comparisons of the United States and Soviet Economies*, Washington 1959, 151–76.

———, "Pressure and Planning in the Soviet Economy," in H. Rosovsky ed., *Industrialization in Two Systems: Essays in Honor of Alexander Gerschenkron*, New York 1966, 266–85.

———, "A Study in Economic Planning," unpublished doctoral dissertation, Harvard Univ. 1961.

M. Manove, "A Model of Administrative Planning and Plan Execution in Soviet-Type Economies," unpublished doctoral dissertation, M.I.T. 1970.

J. M. Montias, *Central Planning in Poland*, New Haven 1962.

———, "On the Consistency and Efficiency of Central Plans," *Rev. Econ. Stud.*, Oct. 1962, *29*, 280–93.

———, "Planning with Material Balances in Soviet-type Economies," *Amer. Econ. Rev.*, Dec. 1959, *49*, 963–85.

G. Sorokin, *Planning in the USSR*, Moscow 1967.

R. Stone, "Consistent Projection in Multi-Sectoral Models," in E. Mailinvaud and M. Bacharach, eds., *Activity Analysis in the Theory of Growth and Planning*, New York 1967, 232–34.

V. G. Treml, "The 1959 Soviet Input-Output Table (as Reconstructed)," in Joint Economic Committee, U.S. Congress, *New Directions in the Soviet Economy*, Part II-A, Washington 1966, 259–70.

———, "New Soviet Interindustry Data," in Joint Economic Committee, U.S. Congress, *Soviet Economic Performance: 1966–1967,* Washington 1968, 145–58.

P. Weitzman, "Planning Consumption in the USSR," unpublished doctoral dissertation, Univ. Mich. 1969.

Transactions Costs and the Demand for Money

By Thomas R. Saving*

Much of the recent research on the demand for money treats money as essentially no different from other goods.[1] Money and goods are treated symmetrically in the budget constraint of consumers and thus, goods and money are assumed to be equally useful in transactions. This situation leads us to derive the demand for money in what is essentially a model of a transactions-costless economy. The lack of consideration of transaction cost and its effect on consumer behavior has led to rather strained explanations of why individuals use or hold money.[2] Such explanations sometimes involve arbitrary payment schedules, balanced portfolios, or perhaps simply a throwing up of the hands and saying that the utility function contains money holdings as an argument.[3] In this paper I attempt to remedy this sad state of affairs by developing the demand for money from a foundation of money's use in reducing transaction cost. The result of this approach is, I believe, both more attractive intuitively and more relevant.

I. A Pure Exchange Economy

The traditional treatment of a pure exchange economy considers individuals with utility functions of the form

$$U = U(x_i); \quad i = 1, \ldots, n \tag{1}$$

where x_i is the rate of consumption of the ith good. Each individual receives an endowment of goods $\langle x_i^o \rangle$ which may not be carried over from period to period. Individuals are then assumed to maximize (1) subject to the constraint

$$\sum_{i=1}^{n} p_i x_i^o = \sum_{i=1}^{n} p_i x_i \tag{2}$$

where x_i^o is the endowment of the ith good, x_i is the rate of consumption of the ith good, and p_i is the price of the ith good in terms of a base good x_b, $1 \leq b \leq n$ so that $p_b = 1$. In this model, transactions costs are zero since any good may be traded as easily as any other and trading time is either irrelevant or it is assumed that trading takes no time.

Let me now introduce transactions time as a choice variable for consumers. That is, I assume that barter transactions take time and that the consumer's time is limited and valuable to him. Thus, rewrite the utility function in (1) as

$$U = U(x_i, l); \quad i = 1, \ldots, n \tag{3}$$

where l is the proportion of a period devoted to leisure. Additionally, let me partition each period of time into two parts: that part spent on leisure (l) and that part spent in barter (B). Thus,

$$l + B = 1 \tag{4}$$

Finally, assume that transaction time is

---

* Professor of economics, Texas A&M University. I am indebted to Karl Brunner, Robert Clower, Allan Meltzer, W. Phillip Gramm, and Karl Asmus for suggestions and criticisms. The research on which this article was based was financed by the National Science Foundation.

[1] Some notable exceptions are Karl Brunner and Allan Meltzer (1970) and Robert Clower.

[2] See, for example, Don Patinkin, ch. 5.

[3] See William Baumol and James Tobin for examples of fixed payments schedules. For an excellent critique of these two approaches, see Brunner and Meltzer (1967).

an increasing function of transactions undertaken:[4]

$$B = G(T),\ \frac{\partial B}{\partial T} > 0 \tag{5}$$

where T represents the level of transactions, defined as the Euclidean length of the vector of the values of excess demands,[5]

$$T = \left\{ \sum_{i=1}^{n} \left[p_i(x_i - x_i^o) \right]^2 \right\}^{1/2} \tag{6}$$

The problem is then the maximization of (3) subject to (2) and (4). Form the following Lagrangian function

$$\begin{aligned} U^* = {} & U(x_i, l) - \lambda_1\left[\sum p_i(x_i - x_i^o) \right] \\ & - \lambda_2[l + B - 1] \end{aligned} \tag{7}$$

Differentiating (5), (6), and (7) and setting the resulting derivatives of U^* equal to zero yields the following set of necessary conditions for the maximization of (7)

$$\begin{aligned} U_i &= \left\{ \lambda_1 + \lambda_2 G' \frac{p_i(x_i - \overset{o}{x_i})}{T} \right\} p_i; \\ & \qquad\qquad i = 1, \ldots, n \\ U_l &= \lambda_2 \end{aligned} \tag{8}$$

so that

$$\frac{U_i}{U_j} = \frac{\lambda_1 + U_l G' \dfrac{p_i(x_i - x_i^o)}{T}}{\lambda_1 + U_l G' \dfrac{p_j(x_j - x_j)}{T}} \frac{p_i}{p_j}; \tag{9}$$

$$i, j = , \ldots, n$$

The result indicated by equations (9) are, as expected, different from the results obtained when transactions costs are assumed to be zero. Perhaps a better idea of the difference can be had by comparing the respective equilibrium marginal rates of substitution with and without transactions costs assuming that the p_i are unaffected by their introduction. This difference is

$$\begin{aligned} & \left(\frac{U_i}{U_j} \right)_T - \left(\frac{U_i}{U_j} \right)_{\sim T} \\ & = \frac{U_l G'\left[p_i(x_i - x_i^o) - p_j(x_j - x_j^o) \right]}{T\lambda_1 + U_l G' p_j(x_j - x_i^o)} \frac{p_i}{p_j} \end{aligned} \tag{10}$$

where $(U_i/U_j)_T$ and $(U_i/U_j)_{\sim T}$ represents the case with and without transactions costs, respectively. Since λ, T, U_l, G', and the p_i are all positive, the sign of (10) depends on the excess demands. In particular, if the consumer is a net supplier of good i, i.e., $(x_i - x_i^o) < 0$, and a net demander of good j, then $(U_i/U_j)_{\sim T} > (U_i/U_j)_T$. In a two-good world this would require an increase in the consumption of good i relative to good j.

This result is intuitively obvious once the effect of the introduction of transactions costs on the prices of the goods is recognized. Essentially, what has happened is that the price received for goods supplied has fallen in the sense that while the same amount of any other good is obtained the transaction now involves a positive expenditure of time. In contrast, the price paid for goods purchased has risen since in addition to the former amount of other goods given up some transactions time must be expended. Thus, it is not surprising that the introduction of transactions costs results in increased consumption of supplied goods (taken as a group) and decreased consumption of demanded goods (again, taken as a group).

---

[4] It can also be assumed that other resources may be substituted for time in conducting transactions. However, for my purpose nothing is gained from this complication.

[5] The Euclidean distance has several advantages. First, it is invariant with respect to changes in the units in which the x_i are measured. Second, it treats goods supplied and demanded in a symmetric fashion. Third, it is mathematically easy to work with. Note that equilibrium in the system requires that $\sum_\pi x_{i\pi} = \sum_\pi x_{i\pi}^o$ for all i, where the π are individuals, which does not imply that $T=0$. In general the equilibrium will be a repetition of the same transactions period after period. It is exactly this repetition which underlies the assertion that in a completely certain world no transactions costs could exist since individuals would simply make a once and for all infinite commitment. However, for my problem the source of the transactions cost function is irrelevant and will not be discussed further. For further elaboration of this point and an interesting discussion of the source of transactions costs, see Brunner and Meltzer (1970).

The effect of introducing transaction cost is illustrated graphically in Figure 1. In the figure the vertical axis represents time per period, which is necessarily unity. The right-hand axis measures units of good 1 and the left-hand axis measures units of good 2. Point A in Figure 1 is the initial endowment of goods and time $(x_1^o, x_2^o, 1)$. Line CC is the budget line for the given prices (p_1, p_2); it represents the relevant boundary of attainable combinations of x_1, x_2, and l under conditions of no transaction cost and non-satiety for x_1, x_2, and l.

The effect of transaction cost is to reduce the height of CC whenever the desired combination of (x_1, x_2) moves away from (x_1^o, x_2^o). This effect is shown in the figure by the movement of the boundary from CC to $C'C'$. Thus allowing for transaction cost reduces the set of attainable combinations of (x_1, x_2, l) by the sum of the two areas (CAC') to the right and left of A. If money is adopted by a community in which individuals are characterized by the situation in Figure 1, this money must have the effect of raising the boundary $C'C'$ (e.g., from $C'C'$ to $C''C''$ in Figure 1).[6]

II. A Money Economy

The essential difference between money and non-money economies is that in a money economy at least one good is universally accepted in transactions so that the number of transactions involved in any trade is at most two; one trade of a supplied item for money and then a trade of money for a desired item. However, the existence of money does not preclude direct barter whenever such barter is less time consuming than two money exchanges. For example, if one desires to exchange two hours work for a transistor radio, he may sell the labor for money and take the money and trade it for a transistor radio, or if he knows someone who will take two hours of labor for a transistor radio then the trade can be made directly. Moreover, in the latter case having to use money may involve both traders in an extra burden. Thus, universal acceptance of money is interpreted to mean that having money is a sufficient condition for exchange but not necessary.

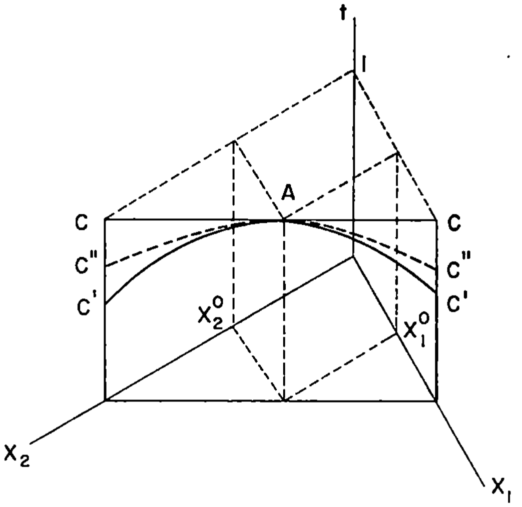


FIGURE 1

One obvious use of money, then, is the reduction of exchange time for the members of the community. In this use, money is used until the marginal cost of using it is equal to the marginal value of time saved through its use. However, money frequently performs other functions such as a convenient store of value; meaning that its storage costs are lower than the storage costs of other goods and services.[7] Thus, the institution of money allows one to avoid storage costs on some commodities, first by making transactions costs lower so that transactions may be engaged in more frequently and secondly, by allowing the separation of the sale of supplied goods from the purchase of goods for consumption.

---

[6] If it were assumed that in addition to time, resources may be used in transactions, then the effect on Figure 1 would be to convert the relevant boundary to a surface rising from line $C'C'$ in the direction of less x_1 or less x_2.

Consider an individual who desires to maximize a utility function of the following form

$$(11) \quad U = U(c_t, l_t); \quad t = 0, \ldots, H$$

where l_t and c_t, respectively, are the rates of consumption of leisure, and goods and services in the period t, and H is the end of the consumers horizon. Assume additionally that the individual faces income and time constraints

$$(12) \quad y_t = H(w_t, t); \; H_{w_t} > 0 \text{ and } H_{w_t w_t} = 0$$

and

$$(13) \qquad l_t + w_t + T_t = 1$$

where y_t is real income in period t, w_t is work time in period t and T_t is the rate of transactions time expenditure in period t.[8] Equation (13) indicates that for each interval the sum of the three rates of spending time must exhaust the interval and it is additionally assumed that the ways to spend time are mutually exclusive.

Any given levels of y_t and c_t imply an aggregate level of transactions. These transactions may be carried out all at once, sporadically, or regularly throughout the interval. Additionally, the transactions may be consummated via direct barter or with a media of exchange. For every distribution of transactions and time path of income and consumption over the interval, there will exist unique levels of average stocks of work services delivered but not collected for, $\bar{y}_t$, consumption goods, $\bar{c}_t$, and real holdings of the various media of exchange, $\bar{m}_t^i$ $(i=1, \ldots, n)$, e.g., currency and demand deposits. Subject to the costs of holding these inventories, consumers will select the distribution of transactions so as to minimize transactions time. I shall assume the following transactions time function[9]

$$(14) \quad T_t = G(y_t, c_t, \bar{y}_t, \bar{c}_t, \bar{m}_t^i); \quad i = 1, \ldots, n$$

where

$$(15) \quad \begin{matrix} G_{y_t} > 0,\ G_{c_t} > 0,\ G_{\bar{y}_t} < 0,\ G_{\bar{c}_t} < 0,\ G_{\bar{m}_t^i} < 0; \\ t = 0, \ldots, H; \quad i = 1, \ldots, n \end{matrix}$$

Since the remainder of this discussion follows directly from assumptions concerning (14) further discussion of it is warranted. First I shall assume that time required for transactions is dependent on both the number and size of the transaction engaged in. Thus, a doubling of income and consumption if handled by doubling the size of each transaction will result in a doubling of $\bar{y}_t$, $\bar{c}_t$, and $\bar{m}_t^i$ and with a resultant increase in transactions costs.[10] The signs of the partial derivatives of (14) (shown as (15)) result from the following: 1) an increase in income must increase transactions time if no change occurs in the other arguments because in this case the only way to sell the increased output is to increase the number of transactions;[11] 2) an increase in consumption increases transactions costs for the same reason as increases in income; 3) an increase in the average stock of produced goods and services implies a reduction in the number of sale transactions (given y_t) and thus a reduction in transactions time; (4) an increase in the average stock of consumed goods and services implies a reduction in the number of purchase transactions (given c_t) and thus a reduction in transactions time; 5) an increase in the average stock of the ith type of money implies (a) larger sale and purchase transactions (given y_t and c_t) and thus fewer of them or (b) that a greater proportion of all transactions are money transactions. Both (a) and (b) above imply a reduction in transaction time.[12]

---

[7] In the context of this paper these storage costs are direct costs since I will deal with a world in which future prices are known to be equal to present prices with certainty.

[8] H_{w_t} and $H_{w_t w_t}$ are the first-and second-order partial derivatives of H. Throughout the remainder of the paper the various functions will be subscripted in a a similar fashion to denote partial differentiation.

[9] Thus, (14) shows the minimum transactions time attainable for each level of the arguments.

[10] Note that if transactions time is independent of the size of transactions but depends only on the number of transactions engaged in, then (14) will be homogeneous of degree zero.

[11] Simply increasing the size of transactions will increase $\bar{y}_t$ and $\bar{m}_t$, which is, of course, not allowed.

Assume additionally that each stock has diminishing returns and that income and consumption have increasing transactions cost so that

$$G_{yy} > 0,\ G_{cc} > 0,\ G_{\bar{y}\bar{y}} > 0,\ G_{\bar{c}\bar{c}} > 0,\ G_{ii} > 0 \quad \text{for all } i \tag{16}$$

and that the only other assets besides the stocks of produced goods and services, consumed goods and services, and the various types of money, are bonds yielding interest rate r per period. Thus, the cost of holding the stocks, $\bar{y}$, $\bar{c}$, $\bar{m}^i$ is the income foregone on bonds that could have been held plus any additional holding costs.[13] Finally, assume that each individual faces a budget constraint of the form shown in equation (17). Note that r is the rate of interest, b is initial real bond holdings, m is initial real money holdings, $\delta_{\bar{y}}$, $\delta_{\bar{c}}$, and the $\delta_{\bar{m}^i}$ are the direct costs of holding inventories of produced goods, consumer goods, and real money, respectively. Additionally, assume that $\delta_{\bar{c}} \geq 0$, $\delta_{\bar{y}} \geq 0$ and $\delta_{\bar{m}^i} < \delta_{\bar{c}}$, $\delta_{\bar{y}}$ for all i, and finally that if for any i, $\delta_{\bar{m}^i} < 0$ then $r \geq |\delta_{\bar{m}^i}|$.[14] The term $r(\bar{y}_t + \bar{c}_t + \sum_i \bar{m}_t^i)$ on the right-hand side of (17) is necessary because the left-hand side of (17) assumes that all of y_t, b, and m is invested at rate of interest r in arriving at the present value. Thus, to the extent that the consumer holds stocks of output, consumption goods, or money, he gives up the return assumed by the left-hand side of (17).

$$\sum_{t=0}^{H}\left(\frac{1}{1+r}\right)^{t} y_t + b + m = \sum_{t=0}^{H}\left(\frac{1}{1+r}\right)^{t} \left[c_t + r\left(\bar{y}_t + \bar{c}_t + \sum_i \bar{m}_t^i\right) + \delta_{\bar{y}_t}\bar{y}_t + \delta_{\bar{c}_t}\bar{c}_t + \sum_i \delta_{\bar{m}_t^i}\bar{m}_t^i\right] \tag{17}$$

Since the problem of interest here is the demand for money, the determinants of the optimal paths of consumption and leisure ($\hat{c}_t$ and $\hat{l}_t$, respectively) may be assumed to have already been chosen; that is, they may be treated as parameters.[15] It can readily be shown that a necessary condition for utility maximization is that for given $\langle \hat{c}_t, \hat{l}_t \rangle$ the levels of $\bar{y}_t$, $\bar{c}_t$, and the $\bar{m}_t^i$ be chosen so that the following implicit form of (17) be a maximum

---

[12] Note that in a certainty world, the minimum $\bar{c}$, $\bar{y}$, $\bar{m}^i$ in each period will be zero.

[13] If the assumption of constant goods prices is dropped and replaced by a known rate of inflation then this rate of inflation may be treated as an additional holding cost of money. In this approach the bonds become fixed in real value so that r represents a real rate of interest. The money rate of interest would then be $r+\rho$, where ρ is the known rate of change in prices. The holding costs of the various forms of money would be $\delta_{\bar{m}^i}+\rho$ provided that all forms of money are suffering the same rate of price level depreciation.

[14] Equilibrium for the economic system with a positive level of capital will require that $r > |\delta_{\bar{m}^i}|$ if $\delta_{\bar{m}^i} < 0$, for all i. Additionally, equilibrium in the banking system will also require the same condition.

[15] This approach can be viewed as a sort of inverse "indirect utility" approach in that instead of the substitution of demand functions into the utility function, I have substituted certain arguments of the utility function into the constraint, i.e., an indirect wealth maximization approach. In this approach the consumer is viewed as a producer and consumer of wealth and given the amount he desires to consume, he maximizes the output of wealth. Thus, I have treated this problem in a way similar to the minimization of cost for given output in the theory of the firm. where even though only one output is the profit maximizing one, it is true that

$$
\begin{aligned}
W^* = {} & \sum_t \left(\frac{1}{1+r}\right)^t y_t + b + m \\
& - \sum_t \left(\frac{1}{1+r}\right)^t \\
& \cdot \left[c_t + r\left(\bar{y}_t + \bar{c}_t + \sum_i \bar{m}_t^i\right) \right. \\
& \left. + \delta_{\bar{y}}\bar{y}_t + \delta_{\bar{c}}\bar{c}_t + \sum_i \delta_{\bar{m}^i}\bar{m}_t^i \right]
\end{aligned} \tag{18}
$$

subject to (12), (13), and (14).[16]

Now substituting (12) and (13) into (18) and (14) yields the following function to be maximized

$$
\begin{aligned}
W^* = {} & \sum_t \left(\frac{1}{1+r}\right)^t H[1 - (\hat{l}_t + T_t), \mathrm{t}] \\
& + b + m - \sum_t \left(\frac{1}{1+r}\right)^t \\
& \cdot \left[\hat{c}_t + r\left(\bar{y}_t + \bar{c}_t + \sum_i \bar{m}_t^i\right) \right. \\
& \left. + \delta_{\bar{y}}\bar{y}_t + \delta_{\bar{c}}\bar{c}_t + \sum_i \delta_{\bar{m}^i} m_t^i \right]
\end{aligned} \tag{19}
$$

subject to

$$
T = G\{H[1 - (\hat{l}_t + T)], \quad \hat{c}_t, \bar{y}_t, \bar{c}_t, \bar{m}_t^i\} \tag{20}
$$

Differentiating (19) subject to (20) and setting the result equal to zero yields the first-order conditions for a maximum of (19)[17]

$$
\begin{aligned}
& -\left(\frac{H_{wt}G_{\bar{y}_t}}{1 + G_{yt}H_{wt}}\right) = (r + \delta_{\bar{y}}) \\
& -\left(\frac{H_{wt}G_{\bar{c}_t}}{1 + G_{yt}H_{wt}}\right) = (r + \delta_{\bar{c}}) \\
& \qquad\qquad t = 0, \ldots, H \\
& -\left(\frac{H_{wt}G_{\bar{m}_t^i}}{1 + G_{yt}H_{wt}}\right) = (r + \delta_{\bar{m}^i}) \\
& \qquad\qquad i = 1, \ldots, n
\end{aligned} \tag{21}
$$

where the terms on the left-hand side of equations (21) are the marginal return to holding average stocks of goods and services produced, goods and services consumed, and money balances, respectively.

---

for any given output costs must be a minimum. In my case this boils down to for any given level of c and l wealth must be a maximum. In particular at the utility maximizing levels of c and l wealth must be a maximum and this maximum is, of course, zero. Essentially my approach says that since (18) must be a maximum for any levels of c and l chosen, (18) must be a maximum for the particular levels of c and l that maximize utility. Treating the problem in this way has the distinct advantage of simplifying the mathematics and of making the demand for money a function of the levels of choices concerning consumption and leisure.

[16] If W^* is not a maximum then some other values of $\bar{y}_t$, $\bar{c}_t$, $\bar{m}_t^i$ must be possible which will result in the left-hand side of (18) being greater than the right-hand side for $\langle \hat{c}_t, \hat{l}_t \rangle$ allowing positive increments to c_t and l_t and therefore an increase in utility. Note that if $(\hat{c}_t, \hat{l}_t)$ is optimal then max $W^* = 0$, but is nonetheless a maximum.

[17] Had I treated the problem as one of utility maximization equations (21) would still have been necessary conditions for a maximum. In addition, the following two conditions would also have been necessary

$$
\begin{aligned}
& U_{c_t} = -\lambda\left(\frac{1}{1+r}\right)^t \left[\frac{1 + H_w(G_y + G_c)}{1 + G_yH_w}\right] \\
& \qquad\qquad \mathrm{t} = 0, \ldots, H \\
& U_{l_t} = -\lambda\left(\frac{1}{1+r}\right)^t \left[\frac{H_w}{1 + G_yH_w}\right]
\end{aligned} \tag{21*}
$$

Moreover, the conditions required for wealth maximization would still be necessary to derive the derivatives of the demand functions presented as equations (22). Essentially, the approach used here derives a wealth function.

$$
W = g(c_t, l_t; r, \delta_{\bar{y}}, \delta_{\bar{c}}, \delta_{\bar{m}^i}) \tag{a}
$$

which shows the maximum wealth for each set of values of the arguments. Utility maximization then requires that the following Lagrangian be a maximum

$$
U^* = U(c_t, l_t) - \lambda[g(c_t, l_t; r, \delta_{\bar{y}}, \delta_{\bar{c}}, \delta_{\bar{m}^i})] \tag{b}
$$

where the function (a) is required to be equal to zero. Thus, the similarity between the usual treatment of profit maximization and the problem considered here is now quite distinct.

The denominators in (21) reflect the fact that the transactions time saved cannot be entirely converted into income for fixed stocks, because increased income requires increased transactions time.

Equations (21) plus the initial conditions (12), (13), and (14) make up, for each t, $(n+5)$ equations in the $(n+5)$ unknowns T_t, w_t, y_t, $\bar{y}_t$, $\bar{c}_t$, $\bar{m}^i_t$ and parameters $\hat{c}_t$, $\hat{l}_t$, r, $\delta_{\bar{y}}$, $\delta_{\bar{c}}$, $\delta_{\bar{m}^i}$. The system, (21), can be solved for each of the variables in terms of the parameters.[18] Let me write the solutions for $\bar{y}_t$, $\bar{c}_t$, $\bar{m}^i_t$ as[19]

$$\hat{\bar{m}}^i_t = \mu^i(\hat{c}_t, \hat{l}_t, r, \delta_{\bar{y}}, \delta_{\bar{c}}, \delta_{\bar{m}}^j, \mathrm{t})$$

$$\hat{\bar{c}} = \gamma(\hat{c}_t, \hat{l}_t, r, \delta_{\bar{y}}, \delta_{\bar{c}}, \delta_{\bar{m}}^j, \mathrm{t}); \tag{22}$$

$$i, j = 1, \ldots, n$$

$$\hat{\bar{y}}_t = \phi(\hat{c}_t, \hat{l}_t, r, \delta_{\bar{y}}, \delta_{\bar{c}}, \delta_{\bar{m}}^j, \mathrm{t})$$

The derivatives of functions (22) can be derived by totally differentiating system (21), (12), (13), and (14). After substitution of (12), (13), and (14) and assuming that the cross-partials involving y and c are zero, the differentiation results in the system of differential equations shown as (23). For expositional convenience I have dropped the time subscripts, and ordered the stocks of the n types of money, produced goods, and consumed goods so that the types of money are denoted as σ_1, σ_2, $\ldots$, σ_n, produced goods as σ_{n+1}, consumed goods σ_{n+2}, $\delta_{\bar{m}^i}$ as δ_i $(i=1, \ldots, n)$, $\delta_{\bar{y}}$ as δ_{n+1}, and $\delta_{\bar{c}}$ as δ_{n+2}.

$$\sum_{j=1}^{n+2} [G_{ij} + (r + \delta_i)(r + \delta_j)G_{yy}]d\sigma_j = (r + \delta_i)\left(\frac{H_wG_{yy}}{1 + H_wG_y}\right)(d\hat{l} + G_c d\hat{c}) - \left(\frac{1 + G_yH_w}{H_w}\right)(dr + d\delta_i); \qquad i = 1, \ldots, n+2 \tag{23}$$

Denote the matrix of coefficients of system (23) as

$$S = [G_{ij} + (r + \delta_i)(r + \delta_j)G_{yy}]; \qquad i, j = 1, \ldots, n+2 \tag{24}$$

then the solution for $d\sigma_j$ is

$$d\sigma_j = \sum_{i=1}^{n+2} (r + \delta_1)\frac{H_wG_{yy}}{1 + H_wG_y}\frac{S_{ij}}{|S|}\cdot(d\hat{l} + G_c d\hat{c}) - \sum_{i=1}^{n+2} \frac{1 + G_yH_w}{H_w}\frac{S_{ij}}{|S|}\cdot(dr + d\delta_i); \tag{25}$$

where $|S|$ is the determinant of S and S_{ij} is the cofactor of S associated with the element common to the ith row and jth column. The second-order conditions for a maximum of (18) require that

$$T = -\frac{H_w}{(1 + H_wG_y)}S \tag{26}$$

be negative definite. Thus, S itself must be positive definite. This information allows the sign of the own partials of the demand functions (22) to be determined, i.e.,

$$\frac{\partial \sigma_j}{\partial \delta_j} = -\frac{1 + G_yH_w}{H_w}\frac{S_{jj}}{|S|} < 0 \tag{27}$$

since $|S|$ and $S_{jj} > 0$ for all j, but leaves the signs of all other partials undetermined.

However, if it is assumed that the effect of a change in the cost of holding the jth stock on the demand for the jth stock is greater than the sum of the effects of changes in all other holdings costs on the demand for the jth stock, i.e.,

---

[18] This follows from the fact that equations (12), (13), (14) are monotonic in T_t, w_t, and y_t, and can be solved uniquely for T_t, w_t, and y_t in terms of the $n+2$ remaining variables. Then these solutions may be substituted into (21). The positive definiteness of S, the Jacobian of the system, then insures the existence of a unique solution of the $n+2$ variables in terms of the parameters.

[19] Note that income and assets do not appear directly in the demand for money function. The reasons for this are twofold. First, income is an endogeneous variable and thus on the same footing as money. However, the income function is assumed fixed (as is the transactions cost function) and will of course influence the demand for money. Second, initial assets affect the demand for money indirectly through c_t, but for our problem c_t is fixed.

$$\left|\frac{\partial\sigma_j}{\partial\delta_j}\right| > \sum_{i\neq j}^{n+2}\left|\frac{\partial\sigma_j}{\partial\delta_i}\right|, \tag{28}$$

then much more can be said.[20] In particular,

$$\frac{\partial\sigma_j}{\partial r} = -\sum_{i=1}^{n+2}\frac{1+G_yH_w}{H_w}\frac{S_{ij}}{|S|} < 0 \qquad \text{for all } j \tag{29}$$

since the sum on the right-hand side of (29) is now dominated by the term $S_{jj}/|S|$ which is positive. In addition, while the effect of a change in desired leisure or consumption on any particular stock is still indeterminate the effect on the sum of all stocks is now determinate and is

$$\begin{aligned}\frac{\partial(\sum\sigma_i)}{\partial\hat{l}} &= \sum_{i=1}^{n+2}(r+\delta_i)\frac{H_wG_{yy}}{1+H_wG_y}\cdot\sum_{j=1}^{n+2}\frac{S_{ij}}{|S|} > 0\\ \frac{\partial(\sum\sigma_i)}{\partial\hat{c}} &= \sum_{i=1}^{n+2}(r+\delta_i)G_c\frac{H_wG_{yy}}{1+H_wG_y}\cdot\sum_{j=1}^{n+2}\frac{S_{ij}}{|S|} > 0\end{aligned} \tag{30}$$

This result follows because (28) implies that $\sum_{j=1}^{n+2}S_{ij}/|S|>0$ for all i so that every term on the right-hand side of equation (30) is positive.

Finally some stronger results can be had if it is assumed that $G_{ij}=0$ for all i and j, $i\neq j$. For in this case it can be established through a series of elementary row and column operations on S that the cofactors of S are[21]

$$\begin{aligned}S_{ij} &= -(r+\delta_i)(r+\delta_j)G_{yy}\prod_{k\neq i,j}^{n+2}G_{kk} < 0; \qquad i\neq j\\ S_{jj} &= \prod_{k\neq j}^{n+2}G_{kk} + G_{yy}\sum_{i\neq j}^{n+2}(r+\delta_i)^2\cdot\prod_{k\neq i,j}^{n+2}G_{kk} > 0\end{aligned} \tag{31}$$

From (25) and (31) it follows that: $\partial\sigma_j/\partial\delta_i>0$ for all $j\neq i$; $\partial\sigma_j/\partial\delta_j<0$, $\partial\sigma_j/\partial\hat{l}>0$, $\partial\sigma_j/\partial\hat{c}>0$ for all j; $\partial\sigma_j/\partial r\gtreqless 0$. Thus, for this case, increases in desired leisure or consumption increase the average level of all stocks, both money and goods. Increases in the storage costs of a stock will decrease the average level of that stock and increase the average levels of all other stocks.[22] Finally, increases in the interest rate may increase, leave unchanged, or decrease a particular average stock depending on that stock's storage costs relative to the storage costs of other stocks. Hence, increases in the rate of interest will not necessarily reduce average holdings of each type of money. However, such increases will reduce the average total value of all stocks held as can be seen from equation (32).

In addition, if the storage costs on all forms of money are less than the storage costs on both produced and consumed goods then an increase in the interest rate will necessarily reduce the average stock of the sum of all monies. This is shown in

---

[20] This is equivalent to assuming that S^{-1} has a dominant diagonal since the $S_{ij}/|S|$ are the elements of S^{-1}.

[21] It can be shown that for any (nxn) matrix $A=(a_{ij})$ where $|a_{ii}|>|a_{ij}|$ for all i and j, $i\neq j$, $a_{ii}>0$ for all i, $a_{ij}=a_{lk}$ for all i, j, l, k, $i\neq j$, $l\neq k$ than a matrix $B=(b_{ij})=A^{-1}$ exists and $b_{ii}>0$ for all i, $b_{ij}<0$ for all i and j, $i\neq j$. Thus, many of the following results will hold if $G_{ij}/(r+\delta_i)(r+\delta_j)=G_{lk}/(r+\delta_l)(r+\delta_k)$ for all i, j, k, l such that $i\neq j$, $k\neq l$. Of course, $G_{ij}=0$ for all i, j, $i\neq j$ is a special case of this condition.

[22] Since a known rate of inflation may be treated as a storage cost of holding money, an increase in the rate of inflation is equivalent to an increase in storage costs of money. Thus, such an increase will reduce the average level of money balances and increase the level of stocks of goods.

$$(32)\quad \frac{\partial(\sum \sigma_j)}{\partial r} = -\left(\frac{1 + G_yH_w}{H_w}\right) \cdot \left\{ \sum_{j=1}^{n+2} \prod_{k\neq j}^{n+2} G_{kk} + G_{yy} \sum_{i,j=1}^{n+2} \cdot [(r+\delta_i) - (r+\delta_j)]^2 \cdot \prod_{k\neq i,j}^{n+2} G_{kk} \right\} < 0$$

equation (33) which is necessarily less than zero if $\delta_j < \delta_{n+1}, \delta_{n+2}$ for all $j \leq n$. Thus, while the value of the entire money portfolio must be negatively related to the interest rate, no individual element in the portfolio need be so related. The reason for this seemingly strange result is that changes in the rate of interest do not leave the relative marginal costs of holding the various stocks unchanged. In particular, consider the marginal cost of holding stocks of *i*th money type relative to the marginal cost of holding stocks of the *j*th money

$$(34)\quad \frac{MC_i}{MC_j} = \frac{r+\delta_i}{r+\delta_j}$$

The derivative of (34) with respect to the interest rate is

$$(35)\quad \frac{\partial\left(\frac{MC_i}{MC_j}\right)}{\partial r} = \frac{\delta_j - \delta_i}{(r+\delta_j)^2}$$

Thus, if $\delta_j < \delta_i$, increases in the interest rate reduce the marginal cost of the *i*th money relative to the *j*th money, and substitution toward the *i*th money form will occur.[23] This substitution effect may be large enough so that an increase in holdings of the *i*th money form result. Note that if the holding costs of money are always less than the holding costs of goods then an increase in interest rates will result in decreases in the price of holding goods relative to any money form.

III. An Alternative Approach

Some additional insight into the nature of the results can be had by viewing the problem as one of the minimization of transactions time for each period τ subject to the condition that (18) equal zero where $\hat{l}_t$, $\hat{w}_t$, $\hat{c}_t$, $(t=0, \ldots, H)$ and the σ_{jt} $(t=0 \ldots, \tau-1, \tau+1, \ldots, H)$ are treated as parameters. Such a minimization of transactions time is a necessary condition

$$(33)\quad \frac{\partial\left(\sum^{n} \sigma_j\right)}{\partial r} = -\left(\frac{1+G_yH_w}{H_w}\right) \sum_{j=1}^{n} \left\{ \prod_{k\neq j}^{n+2} G_{kk} + G_{yy} \sum_{i,j=1}^{n} [(r+\delta_i) - (r+\delta_j)]^2 \cdot \prod_{k\neq i,j}^{n+2} G_{kk} + G_{yy}(r+\delta_{n+1}) [(r+\delta_{n+1}) - (r+\delta_j)] \cdot \prod_{k\neq n+1,j}^{n+2} G_{kk} + G_{yy}(r+\delta_{n+2}) \cdot [(r+\delta_{n+2}) - (r+\delta_j)] \cdot \prod_{k\neq j}^{n+1} G_{kk} \right\} < 0$$

---

[23] In a model with a known rate of inflation ρ, the marginal cost of the *i*th money type relative to the *j*th money type is

$$(34^*)\quad \frac{MC_i}{MC_j} = \frac{r+\rho+\delta_i}{r+\rho+\delta_j}$$

and the derivative with respect to the rate of inflation is

$$\frac{\partial\left(\frac{MC_i}{MC_j}\right)}{\partial \rho} = \frac{\delta_j - \delta_i}{(r+\rho+\delta_j)^2}$$

Thus, if $\delta_j < \delta_i$ an increase in the known rate of inflation reduces the marginal cost of the *i*th money relative to the *j*th money.

for utility maximization since, if transactions time in period τ could be reduced while keeping $\hat{c}_t, \hat{l}_t, \hat{w}_t$ (t=0, . . . , H) and σ_{jt} ($t \neq \tau$) constant, then this additional time could be used to increase $\hat{l}_t$ or $\hat{c}_t$ or both for some or all t and thus increase utility. Viewed in this way the problem is to minimize

$$(36) \quad T_\tau = G(\hat{y}_\tau, \hat{c}_\tau, \sigma_{j\tau}); \quad j = 1, \ldots, n+2$$

subject to

$$(37) \quad \begin{aligned} &(1+r)^\tau \left\{ \sum_{t=0}^{H} \left(\frac{1}{1+r}\right)^t (\hat{y}_t - \hat{c}_t) + b \right. \\ &\left. + m - \sum_{t \neq \tau}^{H} \left(\frac{1}{1+r}\right)^t \sum_{i=1}^{n+2} (r+\delta_i)\sigma_{it} \right\} \\ &= \sum_{i=1}^{n+2} (r+\delta_i)\sigma_{i\tau} \end{aligned}$$

Now form the Lagrangian function

$$(38) \quad T_\tau^* = G - \lambda \left[A - \sum_{i=1}^{n+2} (r+\delta_i)\sigma_{i\tau} \right]$$

where A is the left-hand side of (37).[24] The first-order conditions for the minimization of (38) are

$$(39) \quad G_i = -\lambda(r+\delta_i); \quad i = 1, \ldots, n+2$$

which may be expressed as

$$(40) \quad \frac{G_i}{G_j} = \frac{(r+\delta_i)}{(r+\delta_j)} \quad \text{for all } i \text{ and } j.^{25}$$

From (40) then, the ratios of the marginal costs of holding the various stocks must equal the ratios of their respective marginal transactions time returns.[26] The second-order conditions require that the quadratic form $Q = \sum\sum_{ij} G_{ij} d\sigma_i d\sigma_j$ be positive definite subject to the condition that $\sum_{i=1}^{n+2} (r+\delta_i)\, d\sigma_i = 0$. Thus, the following bordered Hessian must be positive definite.

$$(41) \quad T = \begin{bmatrix} 0 & (r+\delta_j) \\ (r+\delta_i) & G_{ij} \end{bmatrix}; \qquad |T| < 0,\ T_{ii} < 0 \quad \text{for all } i$$

where $|T|$ is the determinant of matrix T and the T_{ii} are the principal minors of matrix T. For each t, equations (39) plus equation (37) comprise (n+3) equations in the (n+3) unknowns (σ_i, λ) and can be solved for the unknowns in terms of the parameters (A, r, δ_i) so that the solution for σ_j may be expressed as[27]

$$(42) \quad \sigma_j = S_i(A, r, \delta_i); \quad i = 1, \ldots, n+2$$

Now the signs of the partial derivatives of equations (42) may be determined by totally differentiating the equation system consisting of (39) and (37) which yields

$$(43) \quad \begin{aligned} &\sum_{j=1}^{n+2} (r+\delta_j) d\sigma_{j\tau} = dA + \sum_{j=1}^{n+2} \sigma_{j\tau}(dr + d\delta_j) \\ &(r+\delta_i)d\lambda + \sum_{j=1}^{n+2} G_{ij} d\sigma_{j\tau} = -\lambda(dr + d\delta_i); \\ &\qquad i = 1, \ldots, n+2 \end{aligned}$$

The matrix of coefficients of (43) is exactly T. Therefore, the general solution of (43) may be written as

$$(44) \quad \begin{aligned} d\sigma_{j\tau} = &\frac{T_{oj}}{|T|} \left[dA + \sum_{i=1}^{n+2} \sigma_{i\tau}(dr + d\delta_i) \right] \\ &- \lambda \sum_{i=1}^{n+2} \frac{T_{ij}}{|T|} (dr + d\delta_i) \end{aligned}$$

---

[24] Note that the level of resources available for transactions is treated as a parameter in its entirety for this problem. In a general utility maximizing model this is, of course, inappropriate.

[25] Since the choice of τ is arbitrary (39) must hold for all τ so that (40) can be written as

$$(40^*) \quad \frac{G_{it}}{G_{jt+k}} = \frac{\left(\frac{1}{1+r}\right)^t (r+\delta_i)}{\left(\frac{1}{1+r}\right)^{t+k} (r+\delta_j)} \quad \text{for all } i, j, \text{ and } k$$

[26] Note the ratios of the marginal transactions time returns equals the ratios of the marginal revenues since the transactions time saved is valued at H_w no matter how it is saved.

[27] The nonsingularity of T, the Jacobian of the system, guarantees the uniqueness of equation (42).

so that the partial derivatives in which I am interested are[28]

$$\frac{\partial \sigma_{jr}}{\partial A} = \frac{T_{oj}}{|T|}$$

$$45)\quad \frac{\partial \sigma_{jr}}{\partial \delta_i} = \sigma_{ir}\frac{T_{oj}}{|T|} - \lambda\frac{T_{ij}}{|T|}; \qquad i, j = 1, \ldots, n+2$$

$$\frac{\partial \sigma_{jr}}{\partial r} = -\sum_{i=1}^{n+2}\sigma_{ir}\frac{T_{oj}}{|T|} - \lambda\sum_{i=1}^{n+2}\frac{T_{ij}}{|T|}$$

Using this approach $\partial\sigma_j/\partial A$ may be viewed as the resource effect, or since the change in A must be the result of a decision to change transactions time in period τ, as a transactions time effect. In this way, $\partial\sigma_j/\partial\delta_i$ may be written as

$$(46)\qquad \frac{\partial \sigma_{jr}}{\partial \delta_i} = \sigma_{ir}\frac{\partial \sigma_j}{\partial A} + M_{ij}$$

where $M_{ij} = -\lambda T_{ij}/|T|$ is interpreted as the substitution effect. That is, M_{ij} is the effect of a change in the costs of holding the ith stock on the average holdings of the jth stock with transactions time constant. Note here that if we assume as earlier that $G_{ii} > 0$ for all i and $G_{ij} = 0$ for $i \neq j$ then

$$T_{oj} = -(r + \delta_j)\prod_{k \neq j}^{n+2} G_{kk} < 0$$

$$(47)\qquad T_{ij} = (r + \delta_i)(r + \delta_j)\prod_{k \neq i,j}^{n+2} G_{kk} > 0$$

$$T_{jj} = \sum_{i \neq j}^{n+2}\frac{(r + \delta_i)^2}{G_{ii}}\prod_{k-1}^{n+2} G_{kk} < 0$$

In this case the signs of the partial derivatives are $\partial\sigma_j/\partial A > 0$ for all j; $\partial\sigma_j/\partial\delta_i > 0$ for all i and j, $i \neq j$; $\partial\sigma_j/\partial\delta_j < 0$ for all j; and $\partial\sigma_j/\partial r \gtreqless 0$; depending on the relation between the costs of holding σ_j relative to the σ_i $(i \neq j)$.

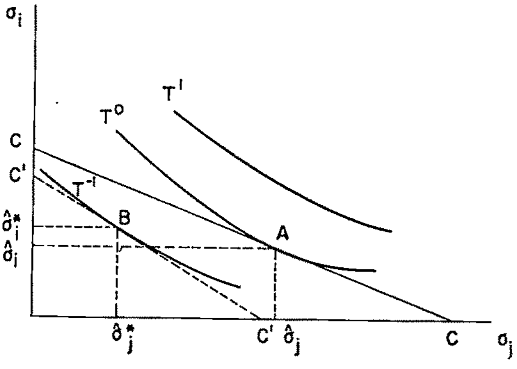


FIGURE 2

Graphically the analysis is depicted in Figure 2 where the vertical axis represents the average stock of σ_i and the horizontal axis represents the average stock of σ_j. The iso-transactions time loci are labeled T^{-1}, T^0, and T^1 and are drawn convex to the origin as required by (41). The constraint is then represented by the linear function with intercepts at

$$\sigma_i = \frac{A - \sum_{k \neq i,j}(r + \delta_k)\sigma_k}{(r + \delta_i)}$$

and

$$\sigma_j = \frac{A - \sum_{k \neq i,j}(r + \delta_k)\sigma_k}{(r + \delta_j)},$$

respectively. The figure is constructed so that $\delta_j < \delta_i$ and the slope of the constraint is accordingly less than one in absolute value. Point A in the figure is the transactions time minimizing combination of σ_i, σ_j given the $\hat{\sigma}_k$ $(k \neq i, j)$. Thus, in this way the vector of optimal stocks is determined. This vector of optimal stocks includes, of course, all the optimal stocks of the various forms of money $\bar{m}^i$ $(i = 1, \ldots, n)$ and thus the portfolio of money forms is determined.

Assume that σ_i and σ_j are both forms of money and allow an increase in the interest rate. As I have shown in (35) such a change will not leave the slope of the constraint unchanged. In fact,

---

[28] Note that the partials with respect to the δ_i and r are compensated partials. That is, it is assumed that the consumer is compensated so that A remains unchanged.

$$
\begin{aligned}
(48)\qquad \frac{\partial\left(\dfrac{d\sigma_i}{d\sigma_j}\right)}{\partial r} &= \frac{\partial\left[\dfrac{(r+\delta_j)}{(r+\delta_i)}\right]}{\partial r} \\
&= -\left[\frac{(r+\delta_i)-(r+\delta_j)}{(r+\delta_i)^2}\right] \\
&= -\frac{(\delta_i-\delta_j)}{(r+\delta_i)^2}
\end{aligned}
$$

so that relative prices change in favor of the more expensive stock when the rate of interest rises. Therefore, it is possible that the substitution effect toward the more expensive money will dominate the resource effect and the increase in the interest rate will increase the stocks of the more expensive money. (Point B in Figure 1 is an example of such a case.) Given a transactions time function that is homogeneous in average stocks, increases in the rate of interest will certainly change the ratios of money stocks in favor of the more expensive stocks.

Thus, if as is sometimes asserted, demand deposits actually have a negative storage cost because of the value of services performed by the bank then increases in the interest rate will increase the currency-deposit ratio.[29] It is interesting to note that the long trend of decreasing currency-deposit ratios has recently reversed itself and that this reversal has coincided with the increasing level of interest rates.[30]

IV. Implications for the Currency-Deposit Ratio

The approach used in this paper can be extended to answer questions concerning the determination of the currency-deposit ratio. In particular, the currency-deposit ratio can be solved for in any of the approaches used by simply limiting the number of money forms to two; currency and demand deposits. In the following, I shall limit my discussion to the approach followed in Section II above.

From the results presented in (12) above the currency-deposit ratio may be written the following way,

$$
(49)\qquad \frac{m_C}{m_D} = \gamma(\hat{c}, \hat{l}, r, \delta_{\bar{y}}, \delta_{\bar{c}}, \delta_C, \delta_D)
$$

where m_C and m_D are real holdings of currency and demand deposits, respectively, δ_C and δ_D are the storage costs of currency and demand deposits, respectively, and the other arguments of (49) are as defined in Section II above.

Now the crucial element in the analysis of the currency-deposit ratio is not (49) but rather the signs of the partial derivatives of (49). These derivatives can be evaluated by first differentiating the quotient m_C/m_D and then assigning the appropriate values to the individual derivatives. First, note that

$$
(50)\qquad \frac{\partial\left(\dfrac{m_C}{m_D}\right)}{\partial x_i} = \frac{m_D\dfrac{\partial m_C}{\partial x_i} - m_C\dfrac{\partial m_D}{\partial x_i}}{m_D^2}
$$

where x_i is the ith argument of (49).

The sign of (50) will depend on the argument chosen and on the assumptions underlying the derivation of the original demand functions. In particular, let me assume that $G_{ij}=0$ for all i and j, $i\neq j$, i.e., the conditions that generated the maximum information concerning the individual derivatives in Section II. In this case, the following solutions for the derivatives of (50) result, and are shown as the system (51). Note that $\alpha=|S|\ m^2_D$, $\beta=(1+G_yH_w)/H_w$ and that G_{CC}, G_{DD} are

---

[29] See Milton Friedman, (p. 42).

[30] As pointed out in fn. 22 an increase in the rate of inflation is similar in its effect to an increase in the rate of interest. Of course the increase in interest rates referred to is the money rate which in my terms is $(r+\rho)$ where ρ is the rate of inflation.

$$\frac{\partial\gamma}{\partial\hat{c}} = \frac{G_{yy}G_{\bar{y}\bar{y}}G_{\bar{c}\bar{c}}G_{c}}{\alpha\beta}\left[m_D(r+\delta_C)G_{DD} - m_C(r+\delta_D)G_{CC}\right] \gtreqless 0$$

$$\frac{\partial\gamma}{\partial\hat{l}} = \frac{G_{yy}G_{\bar{y}\bar{y}}G_{\bar{c}\bar{c}}}{\alpha\beta}\left[m_D(r+\delta_C)G_{DD} - m_C(r+\delta_D)G_{CC}\right] \gtreqless 0$$

$$\begin{aligned}\frac{\partial\gamma}{\partial r} = \frac{\beta}{\alpha}\Big\{&(m_CG_{CC} - m_DG_{DD})G_{\bar{y}\bar{y}}G_{\bar{c}\bar{c}}\\ &+ G_{yy}\left[(r+\delta_{\bar{c}})G_{\bar{y}\bar{y}} + (r+\delta_{\bar{y}})G_{\bar{c}\bar{c}}\right]\left[m_C(\delta_{\bar{c}} - \delta_D)G_{CC} - m_D(\delta_{\bar{c}} - \delta_C)G_{DD}\right]\\ &+ (\delta_C - \delta_D)G_{yy}G_{\bar{c}\bar{c}}G_{\bar{y}\bar{y}}\left[m_C(r+\delta_C) + m_D(r+\delta_D)\right]\Big\} \gtreqless 0\end{aligned}$$

$$\frac{\partial\gamma}{\partial\delta_{\bar{y}}} = \frac{\beta(r+\delta_{\bar{y}})G_{\bar{c}\bar{c}}G_{\bar{y}\bar{y}}}{\alpha}\left[m_D(r+\delta_C)G_{DD} - m_C(r+\delta_D)G_{CC}\right] \gtreqless 0 \tag{51}$$

$$\frac{\partial\gamma}{\partial\delta_{\bar{c}}} = \frac{\beta(r+\delta_{\bar{c}})G_{\bar{y}\bar{y}}G_{\bar{y}\bar{y}}}{\alpha}\left[m_D(r+\delta_C)G_{DD} - m_C(r+\delta_D)G_{CC}\right] \gtreqless 0$$

$$\begin{aligned}\frac{\partial\gamma}{\partial\delta_D} = \frac{\beta}{\alpha}\Big\{&\left[m_D(r+\delta_D)(r+\delta_C)G_{yy} + m_CG_{CC}\right]G_{\bar{y}\bar{y}}G_{\bar{c}\bar{c}}\\ &+ m_CG_{yy}\left[(r+\delta_C)^2G_{\bar{y}\bar{y}}G_{\bar{c}\bar{c}} + (r+\delta_C)^2G_{CC}G_{\bar{y}\bar{y}} + (r+\delta_{\bar{y}})^2G_{CC}G_{\bar{c}\bar{c}}\right]\Big\} > 0\end{aligned}$$

$$\begin{aligned}\frac{\partial\gamma}{\partial\delta_C} = -\frac{\beta}{\alpha}\Big\{&\left[m_C(r+\delta_D)(r+\delta_C)G_{yy} + m_DG_{DD}\right]G_{\bar{y}\bar{y}}G_{\bar{c}\bar{c}}\\ &+ m_DG_{yy}\left[(r+\delta_D)^2G_{\bar{y}\bar{y}}G_{\bar{c}\bar{c}} + (r+\delta_{\bar{c}})^2G_{DD}G_{\bar{y}\bar{y}} + (r+\delta_{\bar{y}})^2G_{DD}G_{\bar{c}\bar{c}}\right]\Big\} < 0\end{aligned}$$

the second own partial derivatives of the transactions time function with respect to currency and demand deposits, respectively.

Note that even in this extreme case only two of the seven derivatives have determinate signs. These two are the derivatives with respect to the storage costs of the two forms of money. As expected, increases in the currency storage cost decrease the currency-deposit ratio and increases in the demand deposit storage cost increase the currency-deposit ratio. However, if I make the additional assumption that changes in the level of consumption leave the currency-deposit ratio unchanged then the remainder of the derivatives are determinate. In this case the signs of the derivatives in (51) become

$$\begin{aligned}&\frac{\partial\gamma}{\partial\hat{c}} = \frac{\partial\gamma}{\partial\hat{l}} = \frac{\partial\gamma}{\partial\delta_{\bar{y}}} = \frac{\partial\gamma}{\partial\delta_{\bar{c}}} = 0\\ &\frac{\partial\gamma}{\partial r} \gtreqless 0 \text{ as } \delta_C \gtreqless \delta_D;\ \frac{\partial\gamma}{\partial\delta_D} > 0;\ \frac{\partial\gamma}{\partial\delta_C} < 0\end{aligned} \tag{52}$$

Thus, the result of assuming that $\partial\gamma/\partial\hat{c}=0$ is that the currency-deposit ratio depends only on the relative price of currency and demand deposits.

V. Conclusion

In this paper I have been able to demonstrate that the theory of demand for money can be derived from utility maximization without making money holdings an argument of the utility function. Essentially, the approach concentrates on the transaction time saving aspect of using money. Thus, the results deal with transactions balances only. The demand for money functions derived have all the properties usually assumed in models where the demand for money is a primitive statement in the model.

In the most general case the only derivative of the demand for any particular money form that is determinate is the own storage cost derivative. Thus, the derivative of the demand for a particular money form with respect to the interest rate is indeterminate. This indeterminancy is due to the fact that the relative costs of using the various money forms is dependent on the interest rate. However, even though it cannot be shown in the general case, that the demand for a particular money form is inversely related to interest rates, it can be shown that the demand for the sum of all stocks held is so related.

Since the model allows for an arbitrary number of money forms, it has direct relevance for questions concerning the optimal currency-deposit ratio. In Section IV the implications of the model for the currency-deposit ratio were derived. It was shown that for somewhat restrictive assumptions the currency-deposit ratio derivatives can all be signed. In particular, the currency-deposit ratio is positively related to the costs of holding demand deposits and inversely related to the costs of holding currency. Moreover, if it is assumed that demand deposits are cheaper to hold than currency, as is often asserted, then the currency-deposit ratio is positively related to both the real and money interest rate. This is an interesting result considering the recent rise in both interest rates and the currency-deposit ratio.

REFERENCES

W. J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Quart. J. Econ.*, Nov. 1952, *66*, 545–56.

K. Brunner and A. Meltzer, "Economies of Scale in Cash Balances Reconsidered," *Quart. J. Econ.*, Aug. 1967, *81*, 422–36.

——— **and** ———, "The Uses of Money: Money in the Theory of Exchange," unpublished 1970.

R. Clower, "A Reconsideration of the Micro-Foundations of Monetary Theory," *Western Econ. J.*, Dec. 1967, *5*, 1–8.

M. Friedman, *The Optimum Quantity of Money*, Chicago 1969.

D. Patinkin, *Money, Interest and Prices*, 2d ed., New York 1965.

J. Tobin, "The Interest Elasticity of the Transactions Demand for Money," *Rev. Econ. Statist.*, Aug. 1956, *38*, 241–47.

COMMUNICATIONS

Interest Rates and the Short-Run Consumption Function

By Warren E. Weber*

In a previous paper in this *Review*, I used the assumption that a representative consumer maximizes utility over a multiperiod horizon to obtain a consumption function which includes the rate of interest as an independent variable. In that study, interest rates were a statistically significant determinant of annual aggregate consumption. In this note, I examine the question whether interest rate changes are also an important determinant of quarterly consumption movements. I will also determine whether the evidence on consumer behavior obtained with the short-run data is consistent with that obtained from the long-run data.

I. Empirical Analysis

This empirical analysis is performed using the consumption function (10) in my previous study. This consumption function is

$$\begin{aligned} C_t &= g^*(\overset{m}{r_t}, e) \\ (1) \quad &\cdot \left\{ W_t + Y_t \sum_{j=0}^{L-e-1} \left[\frac{k_1}{k_3(k_1 + k_2 r_t^m)} \right]^j \right\} \\ &+ \epsilon_t, \end{aligned}$$

where C_t, W_t, and Y_t are aggregate consumption, non-human wealth, and labor income at time t, respectively; r_t^m is a particular market rate of interest at time t; ϵ_t is an independently and identically distributed normal random variable with mean 0 and variance ν^2; β, k_1, k_2, and k_3 are parameters; and $L-e$ is the remaining lifetime of consumers. In (1)

$$g^*(\overset{m}{r_t}, e) = \left[\sum_{j=0}^{L-e-1} (k_1 + k_2 \overset{m}{r_t})^{-\beta\sigma j} \right]^{-1}$$

The parameter $\sigma = 1/(1+\beta)$ is the partial intertemporal elasticity of substitution for consumption. The economic meaning of the other three parameters in the model is best understood if we let γ be consumers' subjective rate of discount for future consumption, ξ be consumers' anticipated rate of growth of labor income, and θ_1 and θ_2 be two parameters which relate the rate of interest at which consumers expect to borrow and lend in all periods (r_t) and a particular market interest rate according to

$$(2) \qquad r_t = \theta_1 + \theta_2 \overset{m}{r_t}$$

Then $k_1 = (1+\gamma)^{-1/\beta}(1+\theta_1)$, $k_2 = (1+\gamma)^{-1/\beta}\theta_2$, and $k_3 = (1+\theta_1)/(1+\xi)$.

The results of the empirical analysis using aggregate quarterly data for the United States for the period 1952–62 are presented in two tables below.[1] In Table 1 we present the results when consumers are assumed to base their interest rate expectations upon a single market interest rate. Five different interest rates are used in the analysis to determine whether the choice of interest rate affects the empirical results.[2]

---

* Assistant professor of economics, Virginia Polytechnic Institute.

[1] The observation period is limited to 1952–62 because the Federal Reserve Board's Flow-of-Funds data used in the construction of both the aggregate non-human wealth and aggregate consumption series are not available for more recent quarters. The data series and description of the method of construction will be sent upon request. All flow data is expressed in terms of quarterly rates and in real terms. I set the remaining lifetime of consumers ($L-e$) equal to 120 quarters. This is the lifetime used in the previous paper.

[2] The five interest rate series are Moody's Aaa and Baa corporate bond yields, the market yields on 3-month bills and 3-to-5 year *U.S.* government taxable securities, and the yield on long-term *U.S.* government bonds. All interest rate series are obtained from the *Supplement to Banking and Monetary Statistics,* Section

TABLE 1—CONSUMPTION FUNCTION ESTIMATES

Parameter	Moody's		U.S. Government Securities		
	Aaa	Baa	3-month bills	3- to 5-year issues	long-term
β	4.573	2.849		4.465	6.541
k_1	1.001	0.9917		1.019	0.9953
k_2	0.1739	0.3749	0.000	0.0146	0.2235
k_3	1.005	0.9949		1.028	0.9980
Sum of sq. residuals	35.61	32.00	45.12	44.75	38.22
Test Stat.	10.4	14.8	0.000	0.362	7.30
Sig. Level	0.001	0.001	0.000	0.7	0.01

When $\theta_2 = k_2 = 0$, changes in the rate of interest will have no effect on short-run consumption decisions. However, if $k_2 > 0 (\theta_2 > 0)$, there is an interest rate effect. Shown in the next to last line of Table 1 are the likelihood ratio test statistics of the null hypothesis that $k_2 = 0$ against the alternative that $k_2 > 0$. In the last line of the table we present the significance levels at which the null hypothesis can be rejected.[3] For both of the Moody's corporate bond rates we can reject the null hypothesis at the 0.001 level of significance. For the long-term government bond rate, we can reject the null hypothesis at the 0.01 level of significance. We cannot reject the null hypothesis for reasonable significance levels for either of the short-term government rates.[4]

Thus, on the basis of our findings in Table 1, we can say that the choice of interest rate does have an important effect on the explanatory power of the model. Changes in corporate bond rates do have effects on short-run consumption, whereas changes in short-term government rates do not. We do not have enough information to make a statement either way about long-term government bond rates.

Now suppose that the explanatory power of an interest rate series is directly related to its reliability as an indicator of the interest rates facing consumers. Then the results indicate that corporate bond rates are more reliable indicators than government bond rates. They also indicate that corporate bond rates are more reliable the higher the default risk of the bond and that government bonds rates are more reliable the longer the maturity of the issue.

We now proceed to allow two interest rates to affect consumers' interest rate expectations by changing equation (2) to

$$r_t = \theta_1 + \theta_2 r_t^m + \theta_3 x_t, \tag{3}$$

where x_t is a second market rate of interest. Using (3) and performing the same manipulations which I used to obtain (10) in my previous study, I obtain a new consumption function which differs from (1) in that wherever $(k_1 + k_2 r_t^m)$ appears in (1), it is now replaced by $(k_1 + k_2 r_t^m + k_4 x_t)$, where $k_4 = (1+\gamma)^{-1/\beta}\theta_3$.

Table 2 presents the results when two interest rates are allowed to affect consumers' interest rate expectations. *Ceteris paribus*, an increase in the second market interest rate will decrease consumers' interest rate expectations. However, since $|\theta_2/\theta_3| = |k_2/k_4|$, θ_3 is always less in absolute value than θ_2, so that an equal increase in both market interest rates will cause consumers' expected interest

---

12. We will use the term "short-term government rates" to refer to the yields on 3-month bills and 3-to-5 year issues. The interest rate data are available upon request.

[3] The level of significance at which the null hypothesis can be rejected is obtained by comparing the value of the test statistic against the percentiles of the chi-square distribution with one degree of freedom. The method of calculation of the test statistics is given in Weber, page 8 and fn. 14. The sum of the squared residuals under the null hypothesis is 45.12.

[4] For 3-month bills the minimum sum of the squared residuals which we obtained is the same as that under the null hypothesis. For this case we cannot empirically distinguish β, k_1, and k_3. Hence, no values for these parameters are reported in the table.

TABLE 2—CONSUMPTION FUNCTION ESTIMATES WITH TWO INTEREST RATES

r_t^m	Moody's Baa				U.S. Gov't 3–5 Year	U.S. Gov't Long-Term	
x_t	Moody's	U.S. Government Securities			U.S. Government		
Parameter	Aaa	3-Month Bills	3- to 5-Year Issues	Long-Term	3-Month	3-Month Bills	3- to 5-Year Issues
β	2.800	2.957	2.597	2.880	4.249	14.95	8.769
k_1	0.9917	0.9919	0.9917	0.9917	1.008	0.9963	1.009
k_2	0.9621	0.4938	0.6800	0.8103	0.2148	0.2679	0.2186
k_3	0.9943	0.9947	0.9943	0.9944	1.013	0.9981	1.015
k_4	−0.6828	−0.2234	−0.3579	−0.5629	−0.1811	−0.1075	−0.1066
Sum of sq. Residuals	28.92	24.44	21.75	28.13	38.12	29.54	28.37
Ω_1	4.73	12.1	17.3	5.95	7.06	11.3	13.1
Sig. Level	0.05	0.001	0.001	0.025	0.01	0.001	0.001
Ω_2	19.6	27.0	32.1	21.5	7.42	18.6	20.4
Sig. Level	0.001	0.001	0.001	0.001	0.025	0.001	0.001

rate to increase. Also note that when two corporate bond rates are used in the analysis, it is the one with the lower default risk which enters with the negative sign and has the smaller effect. When a corporate rate and a government rate are used, it is the government rate which has the negative sign and the smaller effect. When two government rates are used, it is the one with the shorter maturity which has the negative sign and the smaller coefficient. In all cases, it is the market interest rate which had the higher explanatory power alone which has the larger coefficient when combined with a second interest rate. Thus, these results tend to reinforce my conclusions about the relative reliability of the various interest rates as indicators of the interest rate conditions facing consumers.

When two interest rates affect consumers' expectations, changes in market interest rates will not affect short-run consumption only if $\theta_2=\theta_3=0$; i.e., only if $k_2=k_4=0$. There will be an interest rate effect if either θ_2 or θ_3 is not equal to zero; i.e., if either $k_2\neq 0$ or $k_4\neq 0$. The line labelled Ω_2 in Table 2 presents the likelihood ratio test statistics of the null hypothesis that $k_2=k_4=0$ against the alternative hypothesis that either $k_2\neq 0$ or $k_4\neq 0$. The next line shows the levels of significance at which the null hypothesis can be rejected. With the exception of the case in which two short-term government rates are used in the analysis, we can reject the null hypothesis of no interest rate effect at the 0.001 level of significance. These results are much stronger evidence in favor of the contention that interest rates are an important determinant of short-run consumption than are those presented in Table 1.

Does the inclusion of the second interest rate in the analysis significantly increase the explanatory power of the model? The line labelled Ω_1 in Table 2 presents the likelihood ratio test statistics of the null hypothesis that $k_4=0$ $(\theta_3=0)$ against the alternative hypothesis that $k_4\neq 0$ $(\theta_3\neq 0)$. The next line presents the levels of significance at which the null hypothesis can be rejected. With Moody's Baa rate we obtain a significant (at the 0.001 level) increase by including either the yield on 3-month bills or 3-to-5 year issues. However, the null hypothesis is not rejected when either the Aaa rate or the long-term government bond rate is included with the Baa rate. Including either the yield on 3-month bills or 3-to-5 year issues with the long-term government rate significantly increases the model's explanatory power. These results coupled with those presented directly above indicate that two market interest rates provide a better explanation of short-run consumption than does a single market interest rate. These results also lead

Table 3—Comparison of Estimates Using Annual and Quarterly Data

Parameter	Annual Data					Quarterly Data			
	Durand's Basic Yields Years to Maturity							U.S. Gov't Securities	
	1	5	30	Moody's Baa	Av. Annual Yield Time & Saving Deposits In Com. Banks	Moody's Aaa	Moody's Baa	3–5 Year	Long-Term
β	6.288	2.599	1.427	6.675	2.778	4.573	2.849	4.465	6.541
$\sigma=1/(1+\beta)$	0.1372	0.2779	0.4120	0.1303	0.2647	0.1794	0.2598	0.1830	0.1326
γ	−0.3140	−0.1564	−0.0820	0.0274	−0.1446	0.0174	0.0388	−0.0593	0.0658
ξ	−0.1092	−0.0949	−0.0668	−0.0162	−0.083	0.0003	0.0108	−0.0225	0.0070
θ_2	0.1839	0.3500	0.9248	0.5928	0.4295	0.0437	0.0950	0.0036	0.0564

us to conclude that a combination of the Moody's Baa rate and a short-term government rate provides the best indication of the interest rates facing consumers. The expected interest rate calculated from a combination of yields on two different types of bonds provides a more reliable interest rate indicator than either a combination of two corporate bond rates or two government security rates.[5]

II. Comparison of the Long-Run and Short-Run Results

Table 3 presents the estimates of the parameters β, θ_2, γ, and ξ with both annual and quarterly data.[6] The quarterly estimates are obtained from Table 1. The annual estimates are obtained from Tables 1 and 2 in my *Review* article.

Our estimates of the partial intertemporal elasticity of substitution (σ) are similar and uniformly less than 0.5 for both sets of data. These results indicate that consumers do not regard consumption in different quarters to be any better substitutes than they regard consumption in different years. And, regardless of the time period over which the flow of consumption is measured, they regard consumption in different time periods to be poor substitutes.

The two sets of data yield conflicting results concerning the consumers' rate of time preference (γ). The annual data indicate that consumers are impatient, desiring to advance the timing of their consumption. The quarterly data indicate that they are slightly negatively impatient, desiring to slightly postpone the timing of their satisfaction. This difference could arise because next quarter's consumption appears relatively near to this quarter's, so that consumers are almost indifferent whether they have their consumption in this period or the next. However, next year's consumption is relatively far off, so that they may want some of next year's consumption now.

The two sets of data also yield conflicting results about consumers' anticipated rate of growth of labor income (ξ). The annual data indicate that consumers expect labor income to decline over the remainder of their economic lifetimes, whereas the quarterly data indicate that they expect it to increase. This conflict may only be illusory, however, for when we took account of retirement in the study with annual data we found that our estimates of ξ became positive. Since retirement is relatively further off for quarterly consumption decisions, it may not exert the influence on ξ which it apparently does for the annual decisions, so that ξ becomes positive with the quarterly data.

---

[5] I also performed the analysis using the rate of inflation rather than a second market interest rate for x_t. The rationale for including the rate of inflation is that consumers may base their consumption decisions upon real rather than nominal interest rates. Consequently, I expected and found negative signs on k_4. I also found however, that including the rate of inflation only trivially increased the explanatory power of our model, and thus concluded that it was nominal interest rate changes which affected short-run consumption decisions.

[6] We cannot obtain direct estimates of θ_1, θ_2, γ, and ξ unless we are willing to arbitrarily specify a value for one of these parameters. This same problem arose in my previous study; the solution was to note that θ_1 could be interpreted as an interest rate floor or a risk premium which gave a basis upon which to assign a value to this parameter. I chose the value $\theta_1=0.02$ since it was approximately the mean of the average annual yield on time and savings deposits in commercial banks over the period 1930–1965. I set $\theta_1=0.005$ here in order to make the estimates with quarterly data comparable with the estimates from the annual data.

Finally, we find that our estimates of θ_2 are uniformly smaller with the quarterly data than with the annual data. This result is certainly expected and consistent. In the annual study the average effective rate of interest is an annual rate, whereas in the quarterly study it is a quarterly rate. Since a 1 percent increase in a quarterly rate is approximately equal to a 4 percent increase in an annual rate, we would expect changes in the market rate (always an annual rate) to have a smaller effect on consumers' average effective interest rate in the quarterly study.[7] In addition, the two sets of data yield consistent results concerning the magnitude of θ_2 and the type of asset yield upon which consumers are basing their interest rate expectations. We find that θ_2 is larger the longer the maturity of the asset and the greater the default risk of the asset.

III. Conclusion

In my previous study with annual data I found interest rates to be an important determinant of consumption, regardless of the interest rate series used in the empirical analysis. The results obtained with the quarterly data do not allow such a strong conclusion. Here we find that when a single market interest rate is used in the analysis, the choice of interest rate makes a difference in the empirical results. Only corporate bond rates are found to be statistically significant determinants of consumption. It is only when two market interest rates are included in the analysis that we obtain results which are in any way as insensitive to the choice of interest rate series as were the annual results. And even in this case, we find that using only two short-term government rates as explanatory variables does not improve the explanatory power of the model.

All of the evidence presented in this and the previous study supports the hypothesis that interest rates influence consumption. Since all of the empirical evidence in this study which does not support this hypothesis is obtained using exclusively yields on government securities, I feel that the evidence presented for the hypothesis outweighs that presented against it.

REFERENCES

W. E. Weber, "The Effect of Interest Rates on Aggregate Consumption," *Amer. Econ. Rev.*, Sept. 1970, *60*, 591–600.

Board of Governors of the Federal Reserve System, "Money Rates and Securities Markets," *Supplement to Banking and Monetary Statistics, Section 12*, Washington 1966.

---

[7] As this discussion clearly shows, in the annual study θ_2 is a unitless number. However, in the quarterly study it is in terms of years/quarter. Therefore, each of the quarterly indirect estimates of θ_2 are divided by four to get the results presented in Table 3.

Unemployment and Inflation: A Cross-Country Analysis of the Phillips Curve

By DAVID J. SMYTH*

This paper undertakes a cross-section analysis of the relation between unemployment and the rate of inflation using data for eleven countries for the period 1950–60. A markedly non-linear Phillips curve is obtained.[1] The paper also investigates two further hypotheses. First, that a single Phillips curve should be replaced by a family of curves, one for each rate of productivity increase. Secondly, that the average rate of inflation will be higher for a given average level of unemployment, the greater the magnitude of cyclical variations in unemployment. Both these hypotheses are rejected.

The plan of the paper is as follows. Section I describes the data used and the notation. Section II estimates the cross-country Phillips curve. The hypotheses relating to productivity growth and cyclical variations in unemployment are investigated in Sections III and IV, respectively. The paper's conclusions are summarized in Section V.

I

A major difficulty in any study involving comparison of unemployment rates in different countries is that the coverage of unemployment statistics varies widely from country to country. It is important to attempt to convert the estimates to a common basis. Angus Maddison assumes that the census coverage of unemployment is fairly comparable between countries and uses information for census years to make corresponding adjustments for other years.[2] His estimates have been adopted in the present paper.

Eleven countries are used in this study: they are Belgium, Canada, Denmark, France, Germany (Federal Republic), Italy, Netherlands, Norway, Sweden, United Kingdom, and the United States. The availability of Maddison's comparable unemployment series limits the sample to this size;[3] it also dictates the choice of period studied, 1950–60.

The average annual increase in the price index relating to the gross national product over the period 1950–60 is used as the measure of the rate of inflation. (See Maddison, p. 45.) Three measures of productivity growth are used: the average annual rate of growth of output per man-hour, of output per head of population, and of output (see Maddison, pp. 28, 30, and 37). The variance and the coefficient of variation of unemployment over the eleven years 1950–60 are used as measures of the time dispersion of unemployment. These are both calculated from annual data.

The following notation is used.

P = average annual increase in the gross national product price index, 1950–60.

---

* Professor of economics, Claremont Graduate School. I have benefited from discussion with Arthur Butler and James Holmes.

[1] The Phillips curve has been widely interpreted as providing a relationship between the rate of inflation and unemployment (see, for instance, Paul Samuelson and Robert Solow) although in his original article Phillips used change in money wage rates and not change in prices.

[2] Maddison also makes some other adjustments. Details of his procedures and his unemployment series are given in his Appendix E.

[3] Maddison also gives unemployment estimates for Switzerland. However, this country was eliminated from the analysis because of difficulties in reconciling the data used by Maddison and that given in the United Nations *Statistical Yearbook*. The census estimate of the unemployment rate in Switzerland on December 1st, 1950, was 0.42 percent while the registered unemployment series gave a figure of 0.29 percent for November 30th of the same year. (Both figures are from the *Annuaire Statistique de la Suisse*, 1958, and are reported by Maddison (p. 222).) These estimates seem extraordinarily low. The figure for 1950 given in the *United Nations Statistical Yearbook* is 1.8 percent. In view of this wide discrepancy it seemed best to exclude Switzerland from the sample.

U = average unemployment rate, 1950–60.
G_1 = average annual rate of growth of output per man-hour, 1950–60.
G_2 = average annual rate of growth of output per head of population, 1950–60.
G_3 = average annual rate of growth of output, 1950–60.
V = variance of annual unemployment rates, 1950–60.
C = coefficient of variation of unemployment rates, $(=V^{1/2}/U)$, 1950–60.

II

Figure 1 presents the scatter diagram between the average annual price increase and the average unemployment rate. It is apparent that the relationship between P and U is a highly non-linear one. Accordingly, we follow Richard Lipsey and George Perry and make use of non-linear transformations of unemployment.[4] The most satisfactory regression equation obtained is

$$(1) \quad P = \underset{(0.29)}{2.73} + \underset{(1.22)}{9.48}U^{-2} \qquad R^2 = .869$$

The result obtained when U^{-1} is used instead of U^{-2} is inferior and there is little point in using U^{-1} as well as U^{-2} as the two are highly correlated.[5] If U is used instead of U^{-2} we have

$$(2) \quad P = \underset{(0.81)}{6.70} - \underset{(0.198)}{0.655}U \qquad R^2 = .548$$

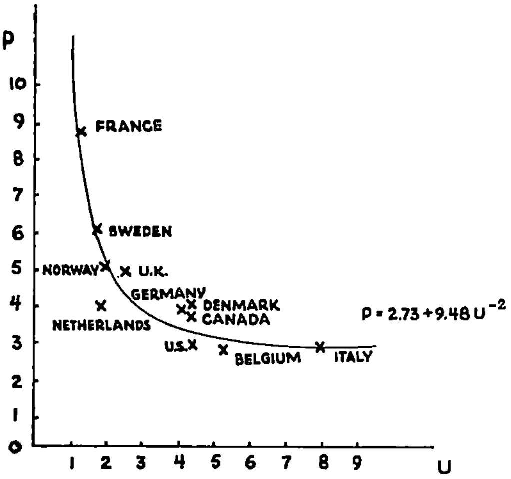


FIGURE 1. AVERAGE RATE OF INFLATION AND AVERAGE UNEMPLOYMENT RATE, 11 COUNTRIES, 1950–60.

This linear relation is markedly inferior to that involving U^{-2}; the F-ratio between the two mean square deviations is 3.46 which is significant at the 5 percent level.

The fit obtained in (1) between the rate of inflation and the unemployment rate is remarkably good for a cross-section analysis.[6] The coefficient of U^{-2} is easily significant at the 0.1 percent level. The regression line given by (1) is plotted in Figure 1. As the unemployment rate falls below about 2 percent the rate of inflation increases markedly and as unemployment rises above about 4 percent little diminution in the rate of inflation results. Given that governments have a trade off between the rate of inflation and the unemployment rate, this suggests that the optimum rate of unemployment is likely to lie between 2 and 4 percent unless one of the policy objectives is weighted relatively heavily compared with the other.[7]

III

Perry has advocated the replacement of a single Phillips curve by a family of curves,

---

[4] Adoption of Lipsey's mathematical form involves no commitment to his derivation of the Phillips curve. James Holmes and I have demonstrated that there is not a unique relationship between the excess demand or supply of labor and unemployment and that consequently the Phillips curve cannot be obtained from Lipsey's model.

[5] Standard errors are given in brackets. When U^{-1} is used instead of U^{-2} the regression obtained is

$$P = \underset{(0.52)}{1.46} + \underset{(1.27)}{7.89}U^{-1} \qquad R^2 = .810$$

The correlation between U^{-1} and U^{-2} is .981. When U^{-1} and U^{-2} are included in the same regression this multicollinearity is reflected in increased standard errors, the estimated regression being

$$P = \underset{(1.01)}{3.32} - \underset{(5.64)}{3.45}U^{-1} + \underset{(6.5)}{13.4}U^{-2} \qquad R^2 = .875$$

[6] It is noteworthy that the biggest deviation from the regression line is for the Netherlands which is substantially negative presumably reflecting the success of its well-known incomes policy during the period.

[7] Converted into terms of the regular *U.S.* unemployment series this range becomes 2.4 to 4.7 percent.

each curve corresponding to a different rate of productivity increase. According to Perry's hypothesis, a rise in the rate of productivity increase will shift the Phillips curve towards the origin; correspondingly a fall in the rate of productivity increase will shift the Phillips curve out away from the origin. On the other hand, Paul Streeten believes that productivity increases are accompanied by equivalent rises in factor payments resulting in no diminution in inflationary pressures.

The present section investigates whether a country's rate of inflation is related to its rate of productivity increase. Three measures of productivity growth are used: G_1, the rate of growth of output per man-hour; G_2, the rate of growth of output per head of population; and G_3, the rate of growth of output. The following regressions are obtained when these three measures of productivity growth are in turn included with U^{-2}.

$$\text{(3)} \quad \underset{(0.72)}{P=2.70}+\underset{(1.32)}{9.47U^{-2}}+\underset{(0.201)}{0.014G_1} \quad R^2=.870$$

$$\text{(4)} \quad \underset{(0.54)}{P=2.70}+\underset{(1.30)}{9.48U^{-2}}+\underset{(0.142)}{0.007G_2} \quad R^2=.870$$

$$\text{(5)} \quad \underset{(0.71)}{P=2.76}+\underset{(1.30)}{9.48U^{-2}}-\underset{(0.150)}{0.009G_3} \quad R^2=.870$$

It is clear from equations (3) to (5) that the inclusion of the productivity variables adds nothing to the explanation of the rate of inflation. A country's rate of inflation appears to be independent of its rate of productivity increase. The cross-country evidence thus supports Streeten's position rather than Perry's.

IV

Lipsey has suggested that if the Phillips curve is non-linear, then the position of the observed Phillips curve will vary with the dispersion of unemployment. Lipsey's argument may be summarized as follows. Suppose that there is more than one labor market, each market having its own identical convex Phillips curve. If unemployment is not the same in each market, then the fitted aggregate curve will lie above the individual market curves. An increase in the dispersion of unemployment will raise the aggregate curve, a decrease will lower it. Using geographical and industrial subaggregates G. C. Archibald and Anthony Thirwall have recently provided support for Lipsey's hypothesis.

We expect to observe the same sort of phenomena as unemployment fluctuates over time. If a country's unemployment rate is not constant but varies cyclically, then the point relating its average price change to its average unemployment rate should lie above the true Phillips curve. Further, for a given average unemployment rate, the point should be higher the greater the dispersion of unemployment about its average over the period under consideration: thus countries that have experienced large cyclical fluctuations during the period 1950–60 should have higher rates of inflation than countries experiencing lesser cyclical fluctuations. To test this hypothesis we calculate two measures of the time dispersion of unemployment from the annual observations for 1950–60, the variance, V, and the coefficient of variation, C, of each country's unemployment series.

Equations (6) and (7) report the effect of including V and C with U^{-2}.

$$\text{(6)} \quad \underset{(0.56)}{P=2.51}+\underset{(1.72)}{10.00U^{-2}}+\underset{(0.221)}{0.099V} \quad R^2=.873$$

$$\text{(7)} \quad \underset{(0.66)}{P=3.32}+\underset{(1.37)}{8.88U^{-2}}-\underset{(2.00)}{1.99C} \quad R^2=.884$$

In neither regression is the dispersion variable significant at the 10 percent level and in (7), it has the wrong sign. We thus conclude that a country's rate of inflation during the period 1950–60 is not influenced by the magnitude of its cyclical fluctuations during the same period and that no support is provided for Lipsey's dispersion of unemployment hypothesis.[8]

---

[8] Various combinations of U, U^{-1}, U^{-2}, G_1, G_2, G_3, V, and C, other than those reported in this paper, were also investigated. In no regression were any of the productivity or dispersion variables significant at the 10 percent level.

V

The conclusions of this paper may be simply stated.

A cross-country analysis of the Phillips curve for the period 1950–60 yielded a downward sloping curve; this curve was markedly convex towards the origin. For a cross-section analysis the fit was remarkably good.

The hypothesis that a single Phillips curve should be replaced by a family of curves, each curve representing a different rate of productivity increase, was tested and rejected; a country's rate of productivity increase had no effect upon its rate of inflation.

The hypothesis that the time dispersion of unemployment influenced a country's rate of inflation was examined. The dispersion variables were not significant, a country's inflation rate thus being independent of the extent of cyclical fluctuations in unemployment.

REFERENCES

G. C. Archibald, "The Phillips Curve and the Distribution of Unemployment," *Amer. Econ. Rev.*, May 1969, *59*, 124–34.

J. M. Holmes and D. J. Smyth, "The Relation Between Unemployment and Excess Demand for Labor: An Examination of the Theory of the Phillips Curve," *Economica*, Aug. 1970, *37*, 311–15.

R. G. Lipsey, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1862–1957: A further Analysis," *Economica*, Feb. 1960, *27*, 1–31.

A. Maddison, *Economic Growth in the West*, New York 1964.

G. L. Perry, *Unemployment, Money Wage Rates, and Inflation*, Cambridge, Mass. 1966.

A. W. Phillips, "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957," *Economica*, Nov. 1958, *25*, 283–99.

P. A. Samuelson and R. M. Solow, "Analytical Aspects of Anti-Inflation Policy," *Amer. Econ. Rev. Proc.*, May 1960, *50*, 177–94.

P. Streeten, "Productivity Inflation," *Kyklos*, 1962, *15*, 723–31.

A. P. Thirwall, "Demand Disequilibrium in the Labour Market and Wage Rate Inflation in the United Kingdom," *Yorkshire Bull. Econ. Soc. Res.*, May 1969, *21*, 66–76.

United Nations, *Statistical Yearbook*, New York 1957.

Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry: An Alternative Approach

By RICHARD D. PORTES*

The purpose of this note is to use the cost function technique to derive the results of the recent article in this *Review* by C. E. Ferguson and Thomas Saving (hereafter F-S). Our interest lies not so much in the results as in the simplicity and power of the technique itself, in contrast to the rather cumbersome conventional apparatus of comparative statics micro-theory.[1] We shall not, therefore, discuss in detail the F-S interpretations of their results, although the technique should give the reader a more intuitive grasp of what is going on.

Following F-S, we assume that the production function $f(x)$ is increasing in each input, strictly concave and twice continuously differentiable on the positive orthant. All equilibrium input vectors are assumed strictly positive. We start from the duality between production and cost functions demonstrated by Shephard, Uzawa, and McFadden; under our assumptions, the production function may be uniquely characterized by its minimum total cost function, defined for all $p>0$ by[2]

$$C = \min_{x \geqq 0} \left\{ \sum p_i x_i \,\middle|\, f(x) = q \right\} = C(p_1, \ldots, p_n; q) \tag{1}$$

The cost function is nondecreasing (increasing at an equilibrium point) and strictly concave in input prices p, and it is increasing and strictly convex in output q. It is twice continuously differentiable in all variables.

We must emphasize here that the duality relationship means that we can equally well begin by assuming these properties of the cost function and omitting any explicit consideration of the production function. A theory of production can take as its "primitive concept" the production function, the technology set, or the cost or profit function. There is now a well-developed theory of the relationships between these concepts (see Shephard (1970)). In any given context, the choice is therefore a matter of suitability to the problem at hand. We would suggest that the cost function is most convenient for a wide variety of problems in the theory of production and demand, from relatively simple questions, like the present one, to much more complex investigations (e.g., Gorman).

We shall write $C' = \partial C/\partial q$, $C_i = \partial C/\partial p_i$, and $C_{ij} = \partial^2 C/\partial p_i \partial p_j$; and we shall use dC/dp_i when we wish to denote that output is being allowed to vary (while other input prices remain constant).

It is a basic result from cost function theory (see Shephard, McKenzie) that under our assumptions,

$$C_i(p_1, \ldots, p_n; q) = x_i(p_1, \ldots, p_n; q), \quad i = 1, \ldots, n, \tag{2}$$

where $x(p; q)$ is the cost-minimizing input vector for q.[3] We shall write $\partial x_i/\partial p_j = C_{ij}$

---

* Assistant professor of economics and international affairs at Princeton University. I wish to acknowledge helpful comments from John Black and the referee.

[1] There is at present no single, general, readily available reference covering the theory of cost and profit functions. This is a pity, since the approach would be quite at home in textbooks on intermediate microtheory. The original reference is R. W. Shephard (1953). The theory was developed further by Hirofumi Uzawa and surveyed by Daniel McFadden. A forthcoming book by Shephard will give by far the most complete development. Applications can be found in Lionel McKenzie, W. M. Gorman (who uses the technique to deal with a problem in the theory of aggregation and gives a detailed, rigorous treatment of profit functions), and Portes (1969).

[2] Our assumptions ensure that the minimum is in fact attained, so that we can write "min" instead of "inf." We use upper-case C to denote the cost *function*, while F-S use lower-case c to denote total cost.

[3] Note that this is F-S (1.8), the essential result from their Appendix Note 1, which underlies the qualitative conclusions of their Section II.

when output is held constant and dx_i/dp_j when output varies.

Since $C(\ )$ is strictly concave in p and strictly convex in q, we have

$$\frac{\partial x_i}{\partial p_i} = C_{ii} < 0, \qquad i = 1, \ldots, n; \qquad C'' > 0 \tag{3}$$

Since the second-order derivatives of $C(\ )$ are continuous, we can reverse the order of differentiation. Noting that C' = marginal cost, we have

$$\frac{\partial}{\partial p_i}(MC) = \frac{\partial C'}{\partial p_i} = C_i' = \frac{\partial x_i}{\partial q}, \qquad i = 1, \ldots, n \tag{4}$$

$$\frac{\partial x_j}{\partial p_i} = C_{ji} = C_{ij} = \frac{\partial x_i}{\partial p_j}, \qquad i, j = 1, \ldots, n \tag{5}$$

The ith input is inferior for given p, over some range of q, if $\partial x_i/\partial q < 0$.[4] We note from (4) that MC at given q varies inversely or directly with p_i according as x_i is inferior or not, i.e., an increase in the price of an inferior input will shift MC downward over the range of outputs for which the input is inferior. F-S obtain this conclusion from their (1.5), which has the same qualitative content as our (4).

All this is unavoidable background, and we have already covered some of F-S. We can now obtain the remaining F-S propositions very quickly indeed. Long-run[5] competitive equilibrium for both firm and industry requires MC = price = AC, i.e., $C' = C/q$. Differentiate with respect to p_i, allowing output to vary:

$$\frac{dC'}{dp_i} = C_i' + C''\frac{dq}{dp_i} = \frac{\partial x_i}{\partial q} + C''\frac{dq}{dp_i},$$

using (4); and

$$\frac{d}{dp_i}\left(\frac{C}{q}\right) = \frac{1}{q^2}\left[q\left(C_i + C'\frac{dq}{dp_i}\right) - C\frac{dq}{dp_i}\right] = \frac{x_i}{q},$$

using (2) and the assumption that $C' = C/q$. Equating these expressions and rearranging, we have

$$\frac{dq}{dp_i} = \frac{x_i}{qC''}\left(1 - \frac{q}{x_i}\cdot\frac{\partial x_i}{\partial q}\right) \tag{6}$$

It is easy to show that this is identical to F-S (21), the key result of Section I of F-S (see the Appendix to this paper, where I show that their η_{ic} equals my $q/x_i \cdot \partial x_i/\partial q$ and their F/cF^* equals my $1/qC''$). Precisely the same qualitative conclusions follow from it: since $C'' > 0$, we have $dq/dp_i \gtrless 0$ according as $\eta_{ic} \lessgtr 1$.

In deriving (6), we found $d(AC)/dp_i = x_i/q$; since by hypothesis $C' = C/q$, we have $d(AC)/dp_i = x_iC'/C$, which is F-S (3.3), the basis of their Appendix Note 3 (p. 783).

Our approach now allows us to give both more precision and more intuitive content to the diagrammatic interpretation in F-S Section III. An input price change shifts the MC curve and the AC curve. In F-S (17) and (21), as in my (6), we can recognize a decomposition of the response of output into a part associated with the shift of the MC curve with constant output price and a part associated with the move along the MC curve caused by the increase of minimum $LRAC$, which determines output price in the long run. We can see this in the following way. Let output price be r. Then we have

$$\frac{dq}{dp_i} = \left(\frac{\partial q}{\partial p_i}\right)_{dr=0} + \frac{\partial q}{\partial r}\,\frac{\partial r}{\partial p_i} \tag{7}$$

Now $\partial q/\partial p_i$ is just the *short-run* response of equilibrium output to the change in input price; the firm adjusts to the shift of the MC curve, taking output price as constant. In

---

[4] For a brief discussion of input inferiority and further references, see Portes (1968).

[5] Note that the distinction between short and long run in F-S, which we follow, is not in terms of whether there are fixed inputs or not—throughout, all inputs are allowed to vary—but rather of whether output price is fixed or allowed to change as the industry adjusts to cost changes.

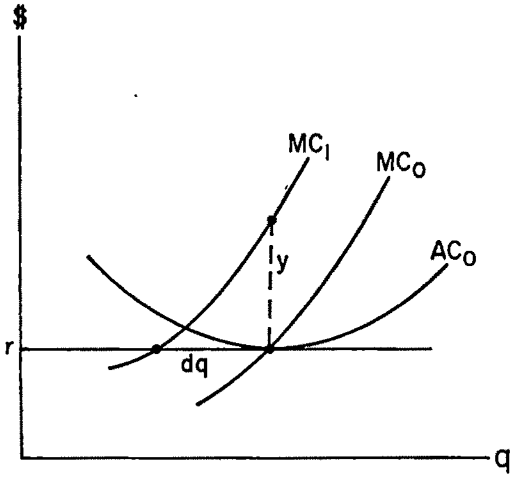


FIGURE 1

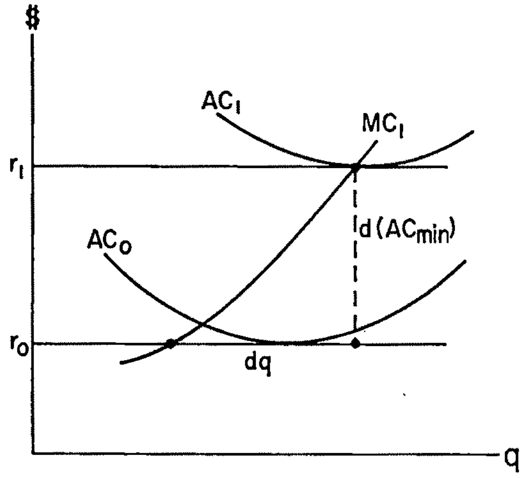


FIGURE 2

the long run, however, output price must change, since the minimum *LRAC* of all firms will have risen due to the input price increase. Consider the second term on the right-hand side of (7): $\partial q/\partial r$ is the change in q per unit movement upward *along* the *MC* curve, the movement upward being measured along the vertical axis; $\partial r/\partial p_i$ is the amount of upward movement per unit increase in p_i, i.e., it is the distance by which minimum *AC* shifts upward as a result of a unit increase in p_i. Thus we can write

$$\frac{\partial q}{\partial r}=\frac{\partial q}{\partial(MC)}=\frac{\partial q}{\partial C'}=\frac{1}{\partial C'/\partial q}=\frac{1}{C''}$$

$$\frac{\partial r}{\partial p_i}=\frac{\partial(AC)}{\partial p_i}=\frac{\partial(C/q)}{\partial p_i}=\frac{C_i}{q}=\frac{x_i}{q},$$

so that

$$\frac{\partial q}{\partial r}\,\frac{\partial r}{\partial p_i}=\frac{x_i}{qC''} \tag{8}$$

We then have

$$\frac{dq}{dp_i}=\left(\frac{\partial q}{\partial p_i}\right)_{dr=0}+\frac{x_i}{qC''} \tag{9}$$

This is equivalent to the decomposition in F-S (17). Indeed, (7) and (9) are just a different way of writing (6), and comparing (9) with (6) gives us the information that

$$\left(\frac{\partial q}{\partial p_i}\right)_{dr=0}=-\frac{1}{C''}\cdot\frac{\partial x_i}{\partial q} \tag{10}$$

A simple geometrical interpretation of (10) and (8) is shown in Figures 1 and 2.

In Figure 1, we show $(\partial q/\partial p_i)_{dr=0}$, the change due to the shift of the *MC* curve. From the geometry, $y=-C''dq$, so $dq=-y/C''$. But y is the vertical shift of the *MC* curve due to the change dp_i, so

$$y=(\partial C'/\partial p_i)dp_i=C'_i\,dp_i=(\partial x_i/\partial q)dp_i$$

Thus $dq=-(1/C'')(\partial x_i/\partial q)dp_i$, which gives (10).

In Figure 2, we show $(\partial q/\partial r)(\partial r/\partial p_i)$, the change due to the move along the *MC* curve. As we found when deriving (6), $d(AC)=(x_i/q)dp_i$. From the geometry, $d(AC)=C''dq$. Thus $dq=d(AC)/C''=(x_i/qC'')dp_i$, which gives (8).[6]

Combining the two figures, we get a somewhat more elaborate version of F-S Figure 1 which is our Figure 3. In each panel, the move from R to S is the first term on the right-hand side of (9), as given in (10) and my Figure 1, while the move from S to T (measured horizontally) is the second term

---

[6] Here $d(AC)$ is the change in *minimum* average cost. The result $d/dp_i(C/q)=x_i/q$ used the assumption that $C'=C/q$.

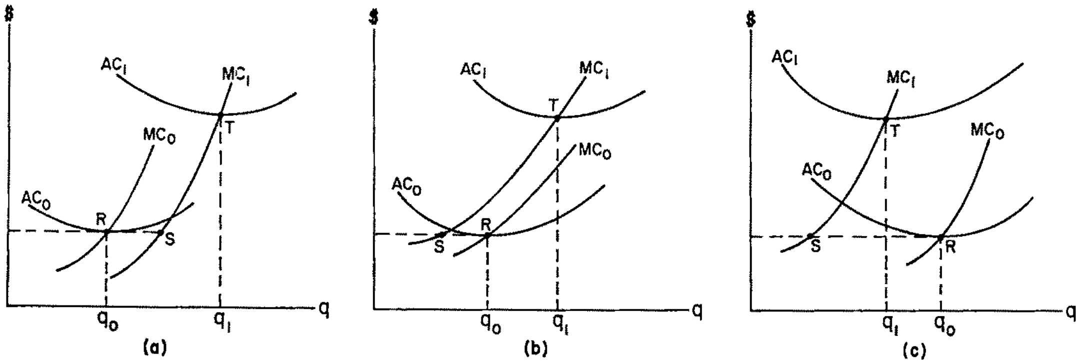


FIGURE 3

on the right-hand side of (9), as given in (8) and my Figure 2. The F-S discussion applies in full, but the underlying situation should now be somewhat clearer. Since $C''>0$, T must always be to the right of S. From (4), we know that when the ith input is inferior ($\eta_{ic}<0$), S will be to the right of R, and thus T will be to the right of R, so output must rise. But if the ith input is not inferior ($\eta_{ic}>0$), S will be to the left of R; if it is superior ($\eta_{ic}>1$), then the horizontal distance between R and S will be greater than that between S and T, so output will fall.

Finally, we briefly discuss F-S (2.7) and (2.10), which give the short-run (output changing, but output price constant) and long-run (output price varying) responses of input demand to a change in input price. Differentiating $x_j=x_j(p;\ q)$ with respect to p_i, we have

$$\text{(11)}\qquad \frac{dx_j}{dp_i}=\frac{\partial x_j}{\partial q}\frac{dq}{dp_i}+\left(\frac{\partial x_j}{\partial p_i}\right)_{dq=0}$$

Here $(\partial x_j/\partial p_i)_{dq=0}$ is just a move along an isoproduct surface (the F-S "substitution effect"). Now the question is simply what happens to q. In the short run, we have $dq/dp_i=(\partial q/\partial p_i)_{dr=0}$, and F-S (2.7) becomes

$$\text{(12)}\qquad \left(\frac{dx_j}{dp_i}\right)_{dr=0}=\frac{\partial x_j}{\partial q}\left(\frac{\partial q}{\partial p_i}\right)_{dr=0}+\left(\frac{\partial x_j}{\partial p_i}\right)_{dq=0}$$

The qualitative information which F-S obtain from (2.7) and (2.8) follows immediately from my (12), using the definition of inferiority together with our (3) and (10).

For F-S (2.10), we are in the long run, so we must substitute (7) into the right-hand side of (11):

$$\text{(13)}\qquad \frac{dx_j}{dp_i}=\frac{\partial x_j}{\partial q}\left[\frac{\partial q}{\partial r}\frac{\partial r}{\partial p_i}+\left(\frac{\partial q}{\partial p_i}\right)_{dr=0}\right]+\left(\frac{\partial x_j}{\partial p_i}\right)_{dq=0}$$

Again, it is easy to show (see the Appendix) that my (13) is precisely equivalent to F-S (2.10). Inspection of (13) together with (3) and (8)–(10) yields all the qualitative information in the F-S discussion of their (2.10), as well as a better intuitive grasp of the relationship between dx_j/dp_i and dq/dp_i in long-run competitive equilibrium.[7,8]

---

[7] John Black has pointed out to me that although the F-S discussion of their (2.10) (my 13) is complete and correct for the individual firm, it leaves open the question of what happens to long-run *industry* demand for an input if its price increases. As F-S say, even if long-run dq/dp_i is positive for the firm, industry output must fall in response to an input price increase, and this is achieved by an exodus of firms from the industry. Since long-run dx_i/dp_i may be positive for the firm, we should answer the corresponding question for input demand. Black has provided the following proof that in the long run, $dX_i/dp_i<0$, where X_i = industry demand for the ith input, Q = industry output, and the industry is composed of homogeneous firms: If $d/dp_i(X_i/Q)<0$, then since $(dQ/dp_i)<0$, we shall have $dX_i/dp_i<0$. Identical

APPENDIX

As stated in the text, the cost function technique is logically independent of and coequal to the production function; that is the import of the basic duality theorem. Thus we started with certain properties of the cost function and deduced all our results directly from these properties, making no reference whatsoever to the production function. It may nevertheless be useful to show explicitly that the equations in the text are in fact equivalent to those of F-S, although the latter are expressed in terms of expenditure elasticities, elasticities of substitution, and the determinants composed of partial derivatives of the production function. It must be stressed, however, that there is no advantage in having the results in this (seemingly more specific) form. In qualitative, comparative statics analysis of the kind involved here, we are only interested in information about signs. As we have seen, this is as readily available from the curvature properties of $C(\ )$ as it is from the signs of the Hessian determinants and their principal minors (the latter being itself equivalent to concavity of the production function). This is simply a consequence of the duality theorem.

We shall demonstrate equivalence for two of the most important pairs of equations, our (6) and F-S (21), and our (13) and F-S (2.10).

By definition, $\eta_{ic}=(\partial x_i/\partial C)\cdot(C/x_i)$. We then have

$$\eta_{ic}=\frac{\partial x_i}{\partial q}\cdot\frac{\partial q}{\partial C}\cdot\frac{C}{x_i}=\frac{\partial x_i}{\partial q}\cdot\frac{C}{C'x_i}$$

Using the fact that in long-run equilibrium, $C'=C/q$, we have

$$\eta_{ic}=\frac{\partial x_i}{\partial q}\,\frac{q}{x_i}. \tag{A.1}$$

From F-S (2.5), if only q changes we have $\partial\lambda/\partial q=-\lambda(F^*/F)$. Now

$$1/\lambda=C',$$

so

$$\begin{aligned} C'' &= -(1/\lambda^2)(\partial\lambda/\partial q)\\ &= (1/\lambda)(F^*/F) = C'(F^*/F)\end{aligned}$$

By hypothesis, $C'=(C/q)$, so we have

$$C''=\frac{C}{q}\,\frac{F^*}{F} \tag{A.2}$$

Substituting (A.1) and (A.2) into F-S (21) gives our (6). Substituting (A.1), (A.2), (8), and (10) into (13), then using F-S (2.1) and (2.2) and $1/\lambda=C'=C/q$, we get F-S (2.10).

REFERENCES

L. R. Bassett and T. E. Borcherding, "The Relationship between Firm Size and Factor Price," *Quart. J. Econ.*, Aug. 1970, *84*, 518–22.

C. E. Ferguson and T. R. Saving, "Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry," *Amer. Econ. Rev.*, Dec. 1969, *59*, 774–83.

W. M. Gorman, "Measuring the Quantities of Fixed Factors," in J. N. Wolfe ed., *Value, Capital and Growth*, Edinburgh 1968.

D. McFadden, "Cost, Revenue and Profit Functions: A Cursory Review," Working Paper 86, Institute of Business and Economic Research, Berkeley, Mar. 1966.

L. McKenzie, "Demand Theory without a Utility Index," *Rev. Econ. Stud.*, 1957, *24*, 185–89.

R. D. Portes, "Input Demand Functions for the Profit-Constrained Sales-Maximizer: Income Effects in the Theory of the Firm," *Economica*, Aug. 1968, *35*, 233–48.

———, "The Enterprise under Central Planning," *Rev. Econ. Stud.*, Apr. 1969, *36*, 197–212.

R. W. Shephard, *Cost and Production Functions*, Princeton 1953.

———, *Theory of Cost and Production Functions*, Princeton 1970.

H. Uzawa, "Duality Principles in the Theory of Cost and Production," *Int. Econ. Rev.*, May 1964, *5*, 216–20.

---

firms implies $X_i/Q=x_i/q$, so $(d/dp_i)(X_i/Q)=(d/dp_i)(x_i/q)=1/q(dx_i/dp_i)-(x_i/q^2)(dq/dp_i)$.

Substituting from (11) and (5) and rearranging, this becomes $(x_i/q^2)[(q/x_i)(\partial x_i/\partial q)-1](dq/dp_i)+C_{ii}/q$. The first term is non-positive, from (6); and the second term is negative, since $C(p,q)$ is strictly concave in p.

[8] A paper by Lowell Bassett and Thomas E. Borcherding discussing some points covered here appeared after this note was written. They relate long-run dq/dp_i to the slope of the firm's expansion path.

Income Taxes and Incentives to Work: Some Additional Empirical Evidence

By D. B. FIELDS AND W. T. STANBURY*

Only very infrequently in the social sciences is it possible to repeat an experiment. This paper is an attempt to repeat an important study whose results were published in this *Review* by George F. Break over a decade ago.[1] It is an attempt to measure the nature and extent of the income (incentive) and substitution (disincentive) effects of the high marginal rates of personal income tax on a group of professionals free to vary their work effort, and more knowledgeable than most individuals about their marginal rate of tax.

In the intervening years, marginal rates in the United Kingdom have declined only slightly and they remain considerably higher than those in both Canada and the United States (see Table 1). While a number of new investigations have been undertaken,[2] none serves to contradict Break's earlier results, and none are as secure methodologically as Break's study. For these reasons it was decided to repeat Break's study.

Richard Musgrave suggests that imposition of an income tax has two opposing effects:

> The disincentive or substitution effect whereby, the price of leisure having fallen, the amount of work declines.
>
> The incentive, or income effect, which moves people to do additional work to restore their previous level of disposable income. Assuming that leisure is a superior good, lower income means a lower demand for leisure. Since theory does not predict the relative strength of these two effects only empirical research can assist in establishing their nature and strength.

I. Methodology[3]

The technique for the selection of the *random* sample of solicitors and chartered accounts in and about London followed very closely the systematic sample employed by Break. The net sample of 285 consisted of 172 in the London sample and 113 in the country sample. Ninety-four of the London sample were solicitors (10 sole proprietors, 84 partners) and 78 were accountants (74 partners and 4 sole proprietors). The country sample had 70 solicitors (12 sole proprietors, 58 partners) and 43 accountants (9 sole proprietors and 34 partners).

A printed questionnaire, patterned after that of Break, was used in all interviews and the questions were always asked in the same order. The questionnaire was divided into sections, as follows:

1. personal reaction of the interviewee to his profession.
2. personal data (age, dependents, fixed financial commitments) and work history (length of time qualified, number of partners, and staff-partner ratio).
3. length of holidays (annual and other)

---

* The interviews for this study were carried out during 1969 by Fields of the faculty of commerce and business administration, University of British Columbia, on a year's study leave. Mr. Stanbury is a doctoral candidate in the economics department, University of California, Berkeley, and has recently joined the faculty of commerce at U.B.C. The study was made possible by the generous financial assistance of the Donner Canadian Foundation. The authors wish to express their sincere thanks to George F. Break of the University of California at Berkeley for his encouragement and advice, and to Sir Thomas Lund of the Law Society and C. Evan-Jones of the Institute of Chartered Accountants in England and Wales for their most helpful cooperation. The London Graduate School of Business Studies kindly provided library and other facilities. Finally our thanks to the hundreds of interviewees for their unfailing courtesy and cooperation.

[1] Sept. 1957, see also Break (June 1957, 1959).

[2] See studies by Rolfe and Furness, L. Buck and S. Shimmin, James Morgan et al., Robin Barlow et al., A. Chatterjee and J. Robinson, *The Graduate Appointments Register*, and C. V. Brown.

[3] A complete statement of the methodology employed is available from the authors upon request.

TABLE 1—COMPARATIVE MARGINAL TAX RATES ON SELECTED EARNED INCOMES IN THE UNITED KINGDOM, CANADA, AND THE UNITED STATES, 1968

Income After Allowable Deductions But Before Personal Exemptions			Marginal Tax Rates on Next Unit of Income Earned (to the nearest percent)					
			Single Person			Married Person with Two Children (under age 11)		
£1	$2.60 Can.	$2.40 U.S.	U.K.	Canada	U.S.	U.K.	Canada	U.S.
£ 400	$ 1,040	$ 960	20	0	14	0	0	0
800	2,080	1,920	41	15	16	20	0	0
2,000	5,200	4,800	41	26	22	41	21	15
4,000	10,400	9 ,600	53	30	28	51	26	19
6,000	15,600	14,400	64	40	36	64	40	22
10,000	26,000	24,000	79	45	50	74	45	32
20,000	52,000	48,000	91	55	60	91	55	50
50,000	130,000	120,000	91	70	70	91	70	52

Notes: Income is "assessed income" after the subtraction of allowable deductions and expenses incurred in earning income. In the case of the United Kingdom, earned income relief and the individual's personal exemption were deducted to obtain "taxable income," and the appropriate marginal tax rate was obtained from the tax schedule. In the case of Canada, the individual personal exemption ($1000) and the minimum standard deduction ($100) were deducted from assessed income to obtain taxable income. In the case of the United States, the individual's personal exemption ($600) and the minimum standard deduction were deducted from assessed income to obtain taxable income.

In this Table we have assumed that £1 in assessed income after deductions allowed by U.K. law, but before personal exemptions, is approximately equivalent to an assessed income of $2.60 (Can.) and $2.40 (U.S.) after deductions allowed by those countries, but before personal exemptions. Break made a similar assumption with the earlier exchange rates.

The 1956 U.K. marginal tax rates were slightly higher than the 1968 rates in each income range.

   and hours of work (currently, previous year, and five years previously.)

4. economic status ("better or worse off") now as compared with five years ago.
5. provisions for retirement (asked only of persons 45 and over).
6. opportunities in past twelve months to accept new work (either personally or for the firm).
7. professional and total income, and estimate of marginal rate of tax.
8. questions relating to the Capital Gains Tax, the special charge and the Selective Employment Tax.
9. questions relating to the incentive or disincentive effects on work effort of high marginal rates of income tax.

By asking questions in this order, it was possible to obtain much useful information before the subject of taxation was introduced. If the interviewee raised the matter of taxation (as a number did, for instance, when asked about their economic status and retirement provisions) this was discussed, and any relevant comments were recorded. However, the questions relating to the impact of taxes were always asked in the prescribed sequence. A very careful analysis of each interview was made to eliminate questionable tax influences. Our emphasis was on actual *changes in behavior* in response to high marginal rates of tax rather than on mere verbalizations.

The overall results of the present study are compared with those of Break (Sept. 1957, Table II, p. 536) in Table 2.

Both studies indicate that a significant proportion of the individuals studied experience or state that they experience incentive or disincentive tax effects, these we have called "gross tax effects." The elimination of questionable tax influences leaves us with those who indicate that their behavior has *changed* as a result of high marginal rates of tax. The subsequent analysis in the paper will deal only with net tax effects.

While Break found that the number of

persons experiencing net tax effects was small (13 percent experienced disincentives and 10 percent incentive effects) but significantly greater than zero, he did not find the disincentive effect to be significantly greater than the incentive effects. However, we found that 18.9 percent of the sample experienced disincentive effects and only 11.2 experienced incentive effects. The difference is statistically significant at the .02 level in a single tail test. The proportion experiencing a disincentive effect has apparently increased over time. It was 13.1 percent in Break's study and 18.9 percent in the current study—the difference is statistically significant at the .03 level. Unfortunately, it is not possible to measure the *net* loss to society in terms of the quantity of labor services.[4]

II. Multiple Tax Effects

At this point in the analysis, an important distinction must be drawn. Throughout his paper (Sept. 1957), Break speaks of the number of "tax effects." Since our study indicated the presence of a substantial number of persons who experienced *multiple* net tax effects (5 of whom experienced *both* incentive and disincentive effects and are counted twice), the number of persons experiencing at least one net tax effect (86) is substantially less than the total number of net tax effects (101). On the other hand, Break records very few multiple net tax effects (5), 4 of which are both incentive and disincentive effects. Consequently he finds that 70 persons record only 71 net tax effects.[5]

TABLE 2—COMPARISON OF THE NUMBER OF PERSONS EXPERIENCING ONE OR MORE TAX EFFECTS

	Distinctive Tax Effects	Incentive Tax Effects
Fields and Stanbury	68/285 = .238	62/275 = .218
	14/285	30/285
	54/285 = .189	32/285 = .112
Break	54/306 = .176	96/306 = .314
	14/306	65/306
	40/306 = .131	31/306 = .101[a]

Note: In each case, questionable tax effects are shown and removed. Thus of our 68 individuals with disincentive effects, 14 are "questionable" leaving 54 with definite disincentive effects.

[a] In his paper, Break indicates 31 net incentive *tax effects*, but in a letter to the authors (Feb. 1970) he indicates that only 30 *persons* are involved. The adjusted ratio, 30/306 = .098, is the one strictly comparable to our results above.

The appearance of multiple tax effects in a single direction is not surprising, i.e., an individual could refuse additional work opportunities and simultaneously shift the work load within the firm so as to do less work. However the existence of both income and substitution effects simultaneously makes one doubt the consistency of such respondents. If we place these respondents in the "questionable" rather than the "net tax effect" category, the change in the results does not affect any of our conclusions.

That we found over one-fifth of the number of persons experiencing tax effects to be experiencing more than one type of tax effect is an important difference between the two studies. The fact that most of these multiple tax effects were disincentive effects reinforces our conclusion that the disincentive effect of continued high marginal rates of tax has increased over time in the United Kingdom.

---

[4] One attempt to do this was carried out in the context of a somewhat similar study by Robin Barlow et al.

While previous empirical evidence indicates that the absolute size of both the disincentive and the incentive effects are small (but significant) and that the two effects about offset each other, one cannot conclude that since total output is little affected there is no cause for concern with the effects of high marginal rates of taxation. As A. R. Prest (p. 265) points out: "The essential point is that we get a different product mix, through shifting of factors trying to minimize tax liabilities, from the one we should get in the absence of high marginal rates (but assuming an equal tax take in some hypothetically innocuous form)."

[5] We found that 9 persons experienced 2 types of disincentive effects and one experienced all 3 types. Only 3 persons experienced 2 types of incentive effect. Because persons experiencing both disincentive and incentive tax effects are counted twice, the 86 persons in our study represent only 81 natural persons. For the same reason the number of natural persons in Break's study is 66.

III. Analysis of the Net Tax Effects

After eliminating all questionable tax influences we are left with 66 disincentive tax effects and 35 incentive tax effects which are shown in Table 3 by *type* of tax effect. Break's 1956 results are shown beside the current results. Not only does the current study show a greater total number of disincentive effects, but it also shows that there has been a considerable shift in the proportion of tax effects recorded in each category. While Break found that 47.5 percent of the net tax disincentive effects reported were in category 1 *Refusal of additional work opportunities*, the current study found only 18.2 percent of disincentive effects in this category. This difference is statistically significant at the .01 level. The current study finds that 45.4 percent of the disincentive consists of a *Shift of work-load within the firm* while Break found only 27.5 percent in this category, the difference is significant only at the .08 level. Break found one-quarter of the disincentive effects in category 2, *Restriction of effort to generate new work,*[6]

TABLE 3—COMPARISON OF TYPES OF NET TAX EFFECTS[a]

Type of Tax Effect[b]		Net Tax Effects Reported			
		Fields & Stanbury		Break	
		Number	Percent	Number	Percent
Disincentive 1	Refusal of additional work opportunities	12	18.2	19	47.5
2	Restriction of effort to generate new work	24	36.4	10	25.0
3	Shift of work load within the firm	30	45.4	11	27.5
Total Disincentive Tax Effects		66	100.0	40	100.0
Incentive 4	Postponed retirement	17	48.6	17	54.8
5	More work on a day to day basis	18	51.4	14	45.2
Total Incentive Tax Effects		35	100.0	31	100.0

[a] Because of the existence of multiple tax effects for one-fifth of those experiencing net tax effects, it is not possible to give this table accurately in terms of *persons* experiencing such tax effects by type.

[b] The following are the definitions of the tax effect categories. Tax effect categories 1, 2, and 3 are disincentive effects, and tax effect categories 4 and 5 are incentive effects.

Category 1. Refusal to accept work offered, or refusal to extend services in other ways (e.g., being more selective in the nature of work accepted, not opening other offices). Interviewee comment: ". . . have refrained from opening a branch office due to taxation; the net return is not worth the risk."

Category 2. Restriction of effort to generate new work. Interviewee comment: ". . . definitely now do not try to generate work, entirely due to tax."

Category 3. Reduction in personal work load through allocation to other staff, or reduction in hours worked. Interviewee comment: ". . . would work longer hours if there was more in it for me at the end after tax."

Category 4. Postponement of retirement, or working longer hours up to date of retirement, due in part to tax burden. Interviewee comment: ". . . would retire tomorrow if financial circumstances permitted; tax is definitely a factor in postponing retirement."

Category 5. Extending hours of work and work effort on a day-to-day basis, partly due to the burden of tax. Interviewee comment: ". . . because of tax burden, have had to take on an undue personal load of work."

---

[6] Break's category 2 was entitled "Reduced incentives to seek new clients," (see Sept. 1957, Table III, p. 539).

Table 4—Comparison of the Number of Accountants and Solicitors, London and Country, Experiencing at Least One Tax Effect

	Accountants			Solicitors			Total
	London	Country	Total	London	Country	Total	
Disincentive	8/78= .103	7/43= .163	15/121= .124	18/94= .191	21/70= .300	39/164= .238	54/285= .189
Incentive	8/78= .103	3/43= .070	11/121= .091	15/94= .160	6/70= .086	21/164= .128	32/285= .112

while in the current study the proportion is 36.4 percent. However this difference is not significantly different statistically. If categories 1 and 2 are grouped together (they are somewhat similar effects, but both are quite different from category 3) we find that the difference between the current results and Break's in the proportion of disincentives in these categories together is significant at the .07 level.

Only slight differences were found between the two studies in the distribution of the two types of incentives, see Table 3.

IV. Relationship of Tax Effects to Other Factors

Accountants and Solicitors

In his 1956 study, Break (Sept. 1957) found that "comparisons of the two professional groups studied showed that solicitors and accountants showed the same reactions to taxation." We find a number of differences, see Table 4.

In terms of net incentive tax effects the difference between accountants and solicitors is not statistically significant. However, the disincentive ratio was significantly greater (.01 level) for solicitors than for accountants.

This change is consistent with other evidence, elicited during the interviews, of a growing disenchantment within the profession on the part of a number of solicitors.[7] Some further evidence that might be germane is that when asked whether they considered themselves better or worse off now, as compared with five years ago, twice as many (proportionately) solicitors as accountants said they were worse off. For all accountants the disincentive ratio was greater than the incentive ratio but the difference is not statistically significant. Solicitors, however, evidenced a significantly higher (.01 level) disincentive tax effect ratio than incentive tax effect ratio.

The Geographical Factor: London and Country

For accountants and solicitors together in the Country group, the disincentive tax effect ratio is .248, but for the London group the comparable ratio is .151. This difference is statistically significant at the .025 level. For the incentive tax effect ratio we find that the London group exceeds the Country group but that the difference is significant only at the .08 level in a single-tail test. Therefore, in the case of the Country professionals, we find that the disincentive tax effect ratio far outweighs the incentive ratio—the difference is highly significant. However, for the London group, the difference is not statistically significant. While Break did not segregate tax effects between solicitors and accountants in his paper, he did report that "tax disincentives tended to occur less frequently among London respondents than among those practising in the smaller communities" (Sept. 1957, p. 544).

---

[7] In reply to the question "If you had a second opportunity would you still choose to become a solicitor (accountant)?", 70.6 percent of accountants replied "Yes" but only 52.8 percent of solicitors did so. When asked about the disadvantages of being a solicitor (accountant) 22.7 percent of accountants replied "none" but only 8.1 percent of solicitors gave this reply. The fact that the Prices and Incomes Board regulates conveyancing fees (an important part of solicitor's income) was frequently cited as a restrictive factor.

TABLE 5—COMPARISON OF THE SENSITIVITY OF PARTNERS AND SOLE PROPRIETORS TO TAXATION

Persons experiencing at least one net tax effect	Partners	Sole Proprietors	Total
Disincentive			
Fields & Stanbury	47/250=.188	7/35=.200	54/285=.189
Break[a]	24/217=.110	16/72=.220	40/289=.138
Incentive			
Fields & Stanbury	24/250=.096	8/35=.229	32/285=.112
Break	21/217=.097	10/72=.139	31/289=.107

[a] Sept. 1957, Table VII.

Partners and Sole Proprietors

Table 5 compares the sensitivity of partners and sole proprietors to taxation. The results are mostly consistent with those of Break, who found "that sole proprietors show a greater sensitivity to both tax incentives and disincentives than do partners" (Sept. 1957, p. 544). Using Break's data we find that sole proprietors had a significantly higher (.02 level) disincentive tax effect ratio than did partners. Making the same comparison in the current study the sole proprietors' tax effect ratio slightly exceeded the partners' ratio but the difference is not statistically significant. In the case of the incentive tax effect ratios, Break's data indicate that while the tax effect ratio for sole proprietors exceeds that of partners, the difference is not statistically significant. However, the current study shows that the sole proprietors' ratio is significantly greater (at the .01 level) than the partners' incentive tax effect ratio.

Break found that "among sole proprietors disincentives tend to exceed incentives, whereas among partners the two types of tax influences are approximately equal" (Sept. 1957, p. 544).

Our results are somewhat different. We find that for partners the disincentive tax effect ratio exceeds the incentive ratio (.188 to .096) and the difference is significant at the .01 level. Among sole proprietors, the current study finds no statistically significant difference in the disincentive tax effect ratio and the incentive ratio. The difference between Break's results and the current one for the disincentive tax effect ratio of partners is significant at the .01 level. In particular, partners are showing a far greater tendency to shift the work load within the firm (disincentive category 3, see Table 3) than they did at the time of the previous study.

The Level of Professional Income and Marginal Rate of Tax

In attempting to compare tax effects on various levels of income, and on incomes earned at different time intervals, it is necessary to take some cognizance of the real value of the monetary unit at the two points of time.[8] For this reason, in Table 6 we used different income classes in our comparison of net disincentive and incentive tax effects than those used by Professor Break in an attempt to compare equivalent levels of *real* income. Because of the range of the income classes we could only approximate comparable levels of real income.

As might be expected when marginal rates of tax are steeply progressive, the results of the study find (consistent with those of

---

[8] Between January 1956 and December 1968 (the period between the two studies) the United Kingdom Index of Retail Prices rose by 46%. Money incomes have risen at least as much. One attempt to measure effective tax rates on *real* income is that by A. J. Merrett and D. A. C. Monk. They use an income multiplier which "represents the *multiple* of the annual rate of inflation in the general price level which the individual had to secure as his increase in before tax income in order that his net of tax income would remain constant in real terms" (p. 95). The formula for the income multiplier is: $m=1-\text{average tax rate}/1-\text{marginal tax rate}$. Using this formula for 1956 and 1968, money incomes would have to rise, on the average, by approximately 53 percent for an individual to have the same real income in 1968 as he enjoyed in 1956.

Break) that the proportion of persons experiencing disincentive effects increases as income increases. Thirty percent of the persons in our sample earning in excess of £7,000 in 1968 experienced a tax disincentive effect, whereas for those earning less than £3,000 there is a disincentive effect experienced by 9 percent. This difference is statistically significant at the .01 level. Break found that tax incentives were rather evenly distributed over the income classes. We find that this pattern has changed, with the incentive effects occurring much more frequently in the middle and lower income classes than in the higher income classes. One notable difference is observed in our £3–4,000 class which is comparable to Break's £2–3,000 class. He found that only 8 percent of persons in this class experienced an incentive effect while we find that 24 percent of persons in the comparable income class experienced incentive effects.

If we group the two lowest income classes in 1968, i.e., under £3,000 and £3–4,000, we find that the incentive tax effect ratio is more than twice that for the two highest income classes taken together. The difference is statistically significant at the .01 level. Grouping those below and above the £4,000 income level, we find that proportion above £4,000 experiencing disincentive effects is significantly greater than the proportion below £4,000 experiencing disincentive effects.

Outside Income

As one would expect, individuals with a significant amount of outside income (over £1,000, equal to U.S. $2400) are far more subject to tax disincentives than those with no outside income.

We found that 39 percent of persons with outside incomes of more than £1,000 experienced at least one net disincentive tax effect while only 15 percent of those reporting no outside income experienced a disincentive effect. The difference is significant at the .01 level.

While 11.3 percent of those with no outside income experienced at least one incentive tax effect and 7.3 percent of those with an outside income of greater than £1,000 experienced an incentive effect, the difference (which is in the direction one would expect) is not statistically significant. The data indicate that of those persons with no outside income the proportion experiencing disincentive tax effects is not significantly different (although it is slightly greater) from the proportion experiencing incentive tax effects. This is not true for those persons with outside incomes exceeding £1,000. Proportionately over five times as many such persons experience disincentive tax effects

TABLE 6—RELATIONSHIP BETWEEN LEVEL OF PROFESSIONAL INCOME AND THE PROPORTION OF PERSONS EXPERIENCING ONE OR MORE NET TAX EFFECTS

		Professional Income in £ Sterling[a]			
Break[b]	1956	<2000	2–3000	3–5000	>5000
Fields & Stanbury	1968	<3000	3–4000	4–7000	>7000
Disincentive					
Break (300 cases)	1956	7/112 = .063	7/63 = .110	10/71 = .141	16/54 = .296
Fields & Stanbury (276 cases)	1968	6/65 = .092	10/58 = .172	20/99 = .202	16/54 = .296
Incentive					
Break (300 cases)	1956	11/112 = .107	5/63 = .079	7/71 = .099	7/54 = .130
Fields & Stanbury (276 cases)	1968	8/65 = .123	13/58 = .241	7/99 = .071	4/54 = .074

[a] does not include "outside" income; see text.

[b] see Sept. 1957, Table VIII, p. 545.

Note: Of our total sample of 285, 9 persons refused to state their income; of Break's total sample of 306 persons, apparently 6 persons refused to state their income.

than experience incentive tax effects, the difference is significant at the .01 level. This result is consistent with our findings that 72 percent of persons experiencing a disincentive are paying surtax, whereas only 41 percent of persons experiencing an incentive effect are paying surtax.

Fixed Commitments

One basic difference is methodology between Break's study and ours occurs in respect of the determination of fixed financial commitments. Break defined commitments as heavy or light *in terms of age and income only* (Sept. 1957, p. 545–46). On the other hand, we asked each respondent the question, "Would you say that your fixed financial commitments are light, moderate or heavy in terms of your income?", and explained that "fixed financial commitments" would include such things as school fees, mortgage payments, pension payments, and capital commitments to the firm.

Break, on the basis of his absolute (and rather arbitrary) definition found that "No respondent with relatively heavy fixed commitments in relation to his income reacted to taxation by contracting his supply of labor; all 26 of the disincentive cases occurred within the group subject only to light commitments" (Sept. 1957, p. 546, Table IX).

Based on the subjective and relative definition (and the interviewee's own estimation) we found *no* statistically significant relationship between the relative burden of fixed commitment and the incidence of incentive or disincentive tax effects.[9]

Age

The results of relating age to the number of persons experiencing tax effects indicate that disincentive effects from taxation are spread fairly evenly throughout the three age groups; under 40, 40 to 50, and over 50. On the other hand, there is a far greater preponderance of incentive effects in the over 50 age bracket than in the other two. We find that 62.4 percent of those persons experiencing at least one incentive tax effect are over age 50 while only 29.7 percent of those experiencing a disincentive effect are in the same age bracket. The difference is statistically significant at the .01 level.

This is explained by the fact that three-quarters of the persons in the over 50 bracket experiencing incentive effects were found to be in category 4, postponed retirement. Obviously, provision for retirement is a serious problem for many professional persons, and not necessarily just those in the lowest income category.[10]

Hours of Work

We found that there are no statistically significant differences between the number of hours worked by persons experiencing tax disincentives, persons experiencing tax incentives, and persons experiencing no tax effects at all. From this it would seem fair to conclude now, as Break did in 1956 that,

> Those who were cutting down their work supplies because of income taxes were still not working less, on the average, than those who were unaffected by taxation. The inference is clear then, that the tax disincentives were concentrated among respondents who were by nature more hardworking and ambitious than the average. [Sept. 1957, p. 546]

V. A Final Note

In survey research of this type one is implicitly asking a hypothetical question—"If taxes were lower (higher) would you work more or work less?" The reply to such a question can only be obtained by introspection. Since the individual hardly operates like a regression model precisely measuring the partial effects of each of a host of variables, while holding *ceteris paribus* it is the task of the researcher to try and weigh the importance of the variables under consideration. Future research seems to lie in the direction of econometrics, but the data for such work may well be derived from more comprehensive survey information. We await the application of such tools.

---

[9] The standard χ^2 test was used.

[10] Two-thirds of the persons experiencing incentive tax effects and over age 50 were found to have professional income in excess of £3,000 (U.S. $7,200).

REFERENCES

R. Barlow, *The Effects of Income Taxation on Work Choices,* Study No. 4 for the Royal Commission on Taxation, Ottawa, 1967.

———, H. E. Brazer, and J. W. Morgan, *Economic Behavior of the Affluent,* Washington, Brookings Institution 1966.

G. F. Break, "Income Taxes and Incentives to Work," *Amer. Econ. Rev.*, Sept. 1957, *47,* 529–49.

———, "Effects of Taxation on Incentives," *British Tax Rev.*, June 1957, 101–13.

———, "Income Tax Rates and Incentives to Work and Invest," *Tax Revision Compendium,* Vol. 3, Committee on Ways and Means, House of Representatives, Washington 1959.

C. V. Brown, "Misconceptions About Income Tax and Incentives," *Scot. J. Polit. Econ.*, Feb. 1968, *15,* 1–21.

L. Buck and S. Shimmin, "Is Taxation a Deterrent?", *Westminster Bank Rev.,* Aug. 1959, 16–19.

A. Chatterjee and J. Robinson, "Effects of Personal Income Tax on Work Effort: A Sample Survey," *Can. Tax J.,* May-June 1969, *17,* 211–20.

A. J. Merrett and D. A. C. Monk, *Inflation, Taxation and Executive Remuneration,* London 1967.

J. W. Morgan, M. H. David, W. J. Cohen, and H. E. Brazer, *Income and Wealth in the United States,* New York 1962.

R. A. Musgrave, *The Theory of Public Finance,* New York 1959, pp. 232–48.

A. R. Prest, *Public Finance,* London 1960.

S. E. Rolfe and G. Furness, "The Impact of Changes in Tax Rates and Methods of Collection on Effort," *Rev. Econ. Statist.*, Nov. 1957, *39,* 394–401.

C. T. Sandford, *Economics of Public Finance,* London 1969, p. 103 cites a survey conducted on behalf of *The Graduate Appointments Register,* April 1967.

Fiscal and Monetary Policy Reconsidered: Comment

By BENT HANSEN*

In a recent issue of this *Review*, Robert Eisner discussed the apparent inefficiency of fiscal policy on the inflationary development in the United States during the last few years. He argued, in particular, that the income tax surcharge could not have been very helpful in dampening the inflationary pressures in 1969 and that expenditure reductions would have been much more efficient. Eisner appraises fiscal policy measures exclusively according to their effects on aggregate final demand per dollar of government revenue or spending. He arrives thus at a ranking with temporary income taxes at the bottom and reductions of public purchases of goods and services at the top. This ranking is identified with a ranking according to anti-inflationary efficiency.

Eisner's reasoning leans heavily upon the life-cycle hypothesis: Since current consumption is largely determined by average expected disposable income, and since the latter cannot possibly change much as the consequence of a surcharge that is expected to last for only a single year, the surcharge could not have had much direct effect on current consumption either. Even after allowance for "normal" multiplier effects (there is little reason to believe that income-earners should perceive the temporary nature of the indirect effects), the impact on the inflationary pressures has, therefore, been small. A similar argument is applied to the effects of the surcharge on investment demand. Because expenditure cuts, even if temporary, affect demand with their full amount and with normal multiplier effects, it stands to reason that their total effects should be substantial. Eisner also mentions the well-known fact that temporary taxes on purchases of (durable) consumer and capital goods have strong direct effects on demand—precisely because they are temporary; several European countries have taken advantage of this circumstance on various occasions. Broadly based indirect taxes (general sales or value added taxes) presumably would be even more efficient for the purpose than the more narrow-based taxes mentioned by Eisner, and measured by demand effect per dollar government revenue, we may have here the most efficient short-term instrument. The possibility of financing the given increase in the war expenditures through sales of government bonds is also discussed. Although Eisner seems to be wrong (even on his own model) in maintaining that at a given money supply, the effects of bond-financed public expenditures cannot, in principle, be anti-inflationary with respect to aggregate demand, he may be right, of course, as a matter of fact.[1]

Thus, there seems to be relatively little to object to in Eisner's reasoning, as far as it goes. But it does not go very far. He is probably justified to blame the advocates of the surcharge for basing their conjectures about its aggregate demand effects upon too crude consumption and investment functions. However, he overlooks the fact that the life-cycle income hypothesis (or, for that matter, the permanent-income hypothesis) is itself a rather narrow theory of household behavior that leaves various reactions out of the picture, and that the effect on aggregate demand is not the only circumstance that matters for a ranking of fiscal (and monetary) policy measures according to anti-inflationary impact.

---

* University of California, Berkeley. My colleagues Robert A. Gordon, Abba P. Lerner, and Earl Rolph commented upon an earlier draft, and the suggestions and critical remarks of an anonymous referee caused me to make some further revisions. I am indebted to all of them. The clerical services provided by the Institute of Business and Economic Research, University of California, Berkeley, are also acknowledged.

[1] I am talking here about financing through sales of ordinary negotiable government bonds. In his concluding remarks, Eisner mentions the possibility of obtaining substantial effects from the sales of nonnegotiable savings bonds with tax or interest benefits.

The life-cycle income hypothesis disregards possible short-term labor supply reactions[2] that may take the place of demand reactions and be significant for the impact on the excess demands in the system and, hence, on the tendencies for money wages and prices to increase. Perhaps more important, from a pratical point of view, are the possible effects on wage claims and profit margins, that is, the possible "cost-inflationary" effects of the surcharge and other fiscal and monetary policy measures. Eisner does not consider such effects either. A ranking of policy measures according to their effects on cost inflation may be entirely different from a ranking according to the effects on aggregate demand or excess demands in the system.

Taking into account short-term labor supply reactions, Eisner may be wrong to maintain that the surcharge has only a minor anti-inflationary impact through its effects on the demand-supply situation, although the possible labor supply reactions probably are too weak to upset his ranking of the policy measures. Cost-inflationary considerations, on the other hand, strengthen his conclusion that a reduction of other defense expenditures would have been one of the most effective ways of curbing the inflationary tendencies arising from the Vietnam War.

My first point is thus that a temporary income tax such as the surcharge will influence labor supply and may be followed by a temporary increase in the supply of labor from households affected by the surcharge. This reaction requires a relatively strong income effect and weak substitution effect on leisure. That the substitution effect should be small in the short run seems reasonable, and at least for families with institutionalized savings—and typically these are wage and salary earners—it seems likely that the short-term income effect should be strong. Considering the relatively small size of the surcharge (measured on disposable income), family members normally not in the labor force (married women, say) may not be induced, to any significant extent, to enter the labor force temporarily. But income-earners already in jobs might take extra jobs, work overtime, and so forth, and unemployed family members already looking for work might cut short their search efforts and accept jobs not considered sufficiently attractive otherwise.

To the extent that the labor supply actually increases, the anti-inflationary effects are obvious. The economy will find itself with an easier labor market and a weaker excess demand pressure on money wages. The rate of increase of money wages should thus tend to be dampened, and if price markups are triggered mainly by money wage increases (at given markup ratios), there will be corresponding dampening effects on price inflation. Moreover, if the households actually succeed in increasing their employment and, hence, their income before tax, production and supply of commodities and services will increase at an unchanged demand (assuming, in line with Eisner's reasoning, that consumption is approximately constant in any case). If the price markup ratios are flexible and influenced by excess demand for commodities, price inflation will be dampened further in this way.

Of course, Eisner may brush these arguments aside in the same way that he deals with the possible effects of general bond sales and changes in money supply—namely, by putting the burden of empirical proof on those who suggest that such effects may be strong. This, perhaps, is as it should be. Existing empirical labor supply studies do not throw much light on the concrete problem at hand, however. Short-term changes in labor supply are certainly strong enough to have significant effect on the labor market situation, but seem to be induced by changes in unemployment rather than wages. In the present case we are confronted, however, with temporary changes in disposable income and their effects may differ from those of permanent wage changes. Be this as it may, Eisner's analysis remains incomplete as long as he does not consider labor supply effects.

---

[2] This point has been made earlier in a more general way. In a survey of consumption theory Daniel Suits says:

> . . . during periods of prosperity and full employment, many spending units are free to offer more or less labor services and hence, within limits to determine their own incomes. The behavior of the spending unit results in the selection of both a volume of consumption expenditures *and* a level of income. . . . " Suits concludes his discussion by stating: "To the extent that steps (e.g. tax reduction) taken to raise incomes result in moderating pressures on the spending unit to produce, they may be offset by reduced earning effort. [p. 23]

My second point is that fiscal policy measures may have effects on wage claims and profit margins and thus have cost-inflationary effects that should be considered when discussing the optimal policy mix. It is convenient to start out from the empirically, relatively well-established hypothesis that the rate of change of money wage rates is partly determined by the rate of change of prices. This relationship implies that an increase in indirect consumer taxation, imposed to diminish consumer demand and thus inflation, will tend to have cost-inflationary effects that might not have arisen had the same demand curb been accomplished by other means. What is gained on the demand side may thus be lost on the cost side. Temporary, broadly based consumer taxes which, as already mentioned, may have the strongest demand effects of all fiscal policy measures, may also have the strongest cost-inflationary effects. It might be argued that a cost inflation of this type can be, at most, of limited size and duration, and if the indirect taxes are temporary, it will be reversed sooner or later. But this is beside the point. In relation to the balance of payments, for instance, even a temporary cost inflation may be dangerous; once the exchange reserves are lost, they are lost forever. *Ceteris paribus* with respect to the demand-supply situation, we shall always be interested, therefore, in minimizing possible cost-inflationary effects, whether temporary or permanent.

Thus, granted that indirect taxes imposed to curb consumer demand may have cost-inflationary effects, it is difficult to see why this should not apply to any kind of taxation that aims at curbing the consumer demand of wage-earners. Union policies may be directed toward securing a certain standard of living rather than certain money, or real wages before taxes.[3] And when wage claims arise directly from the rank and file or from the shop floor—and recent labor market events can leave no doubt that wage claims sometimes do arise in this way—they are probably generated by a strong dissatisfaction of wage-earners with respect to their *general* economic position (disregarding purely political motives). Whether the economic position of the man on the shop floor has been eroded through increasing prices, the surcharge, an increase in payroll taxes, or even an increase in loan rates may matter little to him. What does matter is the increasing difficulty his family has in making ends meet—for whatever reason. Wage claims may be triggered by anything that makes wage-earners feel worse off. If we insist upon this, we should admit, however, that anything that makes wage-earners feel better off may dampen wage claims. Public expenditures made to improve public services to wage-earners could conceivably have such an effect, particularly if they directly and obviously lead to savings for wage-earner households. Improved public health services that directly reduce the household's need for private health service might be a case in point.

The reasoning in the last paragraph applies, perhaps, mainly to spontaneous wage claims and wage actions arising directly from the shop floor.[4] We do, however, live in a time when spontaneous labor market actions seem to become more and more widespread and important. They are not just a "British disease," and policy makers may have to consider them in any country in any context. Economic theory should also try to live up to the situation.

To complete the discussion of cost-inflationary effects, we should, in principle, also take into account effects of business taxes on prices, and effects of public services on private production costs (improved road services reduce private transport costs, and so forth). There is much disagreement about the possibilities for business to shift income taxes on to prices, but at least for corporate business cost-inflation may arise in this way. The effects of public services on private costs cannot be disputed, but from a short-run point of view, this is probably a secondary matter. Finally, it might be argued that since profit margins in contracts about delivery of military equipment tend to be excessive, a cut in defense expenditures might be cost-deflationary through spill-over effect on profit margins in "civil" business.

---

[3] This point has also been made by Brennan and Auld.

[4] The recent Swedish civil servant strike was a union initiated attempt to restore relative disposable salaries as a reaction to the governments' equalization policies.

To conclude, it would seem that the ranking of *temporary* fiscal policy measures with respect to their cost-inflationary effects differs radically from their ranking with respect to demand (or excess demand) effects. On the latter criterion, the ranking, according to effect on demand per dollar revenue or expenditure, may be: income taxes, reduction of purchase for civil or defense purposes, broadly based taxes on purchases of goods and services. On the cost-inflationary criterion, I would tentatively suggest the following ranking: reduction of defense expenditures, taxes on investment, reduction of "useful" civil expenditures, income taxes, and consumer goods taxes. The reasons for this ranking have partly been given above: reduction of defense expenditure may have a positive cost-deflationary effect, and investment taxes do not affect consumer goods prices at present (we are only interested in short-term effects). Reduction of useful civil expenditure may be felt by wage earners through the need for private expenditures. At the upper end we have, of course, income taxes and, at the top, consumer goods taxes.

Granted then, that at given demand-supply effects of alternative fiscal and monetary policy measures, inflation should be curbed by measures that have the smallest adverse effects on wage claims and profit margins, there can be little doubt that a cut in other defense expenditures, given the escalation of the Vietnam War expenditures, must rank high among the conceivable measures that could have served to improve the balance between demand and supply. A surcharge or consumption tax financed war that nobody wants to pay for could, on the other hand, through continuous wage claims and price markups, even lead to one of those permanent inflations where the participants in the social cake party together claim more than one hundred percent of the cake, and each participant administers a parameter that can restore his share each time another participant has succeeded in cutting it down (see George Akerlof). Such consequences tend to threaten in particular when the war also leads to a loosening of the social structure, thus making spontaneous labor actions both more frequent and more successful.

REFERENCES

G. A. Akerlof, "Relative Wages and the Rate of Inflation," *Quart. J. Econ.*, Aug. 1969, *83*, 353–74.

G. Brennan and D. A. L. Auld, "The Tax Cut as an Anti-Inflationary Measure," *Econ. Rec.*, Dec. 1968, *44*, 520–25.

R. Eisner, "Fiscal and Monetary Policy Reconsidered," *Amer. Econ. Rev.*, Dec. 1969, *59*, 897–905.

D. B. Suits, "The Determinants of Consumer Expenditure: A Review of Present Knowledge," in D. B. Suits et al., eds., *Impacts of Monetary Policy: Commission on Money and Credit*, Englewood Cliffs 1963.

Fiscal and Monetary Policy Reconsidered: Comment

By JOHN H. HOTSON*

I accept Robert Eisner's thesis that " . . . the tax surcharge should never, on basic theoretical grounds, have been considered an effective anti-inflationary device and that, given a sufficiently excessive rate of government spending, there is little that any meaningful monetary policy can do to stop inflation" (p. 898). I also share his concern that the failures of current policies will turn our fates back to "know-nothings." I am critical not so much of what Eisner says, as of what he omits.

Is the Johnson administration's desertion of the "Guideposts" in the presence of the inflationary enemy in 1966 of no value in explaining our quickened inflation since then? If the administration had escalated that particular effort, had rallied public opinion, had acquired ultimate legal sanctions against noncompliance, would not the inflation have been lessened? George Perry's findings that the guideposts had a "significant" effect on the pace of wage changes (1967, p. 903) appear to have survived all attacks to date.[1] Perry found that during 1965 and 1966 the guideposts were reducing wage increases by about 2 percent below what would otherwise be obtained. If this level of effectiveness had been maintained, wage payments would have been reduced by about $10 billion in 1968.[2] This is as large as the impact upon demand which was expected from the $10 billion surtax, an impact which Eisner argues did not materialize because the surtax did not change personal and corporate estimates of permanent income (p. 898). Moreover, wage restraint holds down the cost level; in contrast a policy which allows excessive income gains, and then tries to tax them away involves us with barn doors and stolen horses.

The "Keynesian" economists would be less discomforted by know-nothings, in my opinion, if they themselves had been closer students of Keynes. No better place for a fresh start for arriving at correct analysis can be found than in the good book *General Theory*. It is all there: the Phillips curve,[3] the guidepost prescription,[4] and Keynes' Theory of the Price Level of which so few Keynesians appear to be even aware.

> In a single industry its particular price-level depends partly on the rate of remuneration of the factors of production which enter into its marginal cost, and partly on the scale of output. *There is no reason to modify this conclusion when we pass to industry as a whole.* [p. 294] (emphasis added)

Keynes railed against the dichotomic economics of the "classicists" in which no bridge joins the theory of relative prices with the theory of the price level. The first was made to depend on the "homely but intelligible" concepts of micro-theory, while the second depended on the quantity of money. Keynes sought to " . . . escape from this double life and to bring the theory of prices as a whole back to close contact with the theory of value" (p. 293). But in vain! Keynes disciples perpetuate the classical schizophrenia. Chapter 21, "The Theory of Prices," might as well not have been written, so little use is made of it by Keynesians.

---

* Professor of economics, University of Waterloo.

[1] See comments of Paul Anderson, Michael Wachter, Adrian Throop and Perry's reply, 1969. Perry demonstrates in his reply that wage change behavior since 1966 are "just as the guidepost hypothesis would predict. This behavior is not predicted by any alternative hypothesis" (p. 369).

[2] Compensation of employees totaled $513.6 billion in 1968. (See Council of Economic Advisors, p. 241.)

[3] That there is a "trade off" between wages and employment is clear from the following:

> That the wage-unit may tend to rise before full employment has been reached, requires little comment or explanation. . . . there is naturally for all groups a pressure in this direction, which entrepreneurs will be more ready to meet when they are doing better business. For this reason a proportion of any increase in effective demand is likely to be absorbed in satisfying the upward tendency of the wage-unit. [p. 301]

[4] Almost the entire 1962 guidepost policy is contained in and advocated in the following:

> . . . I am now of the opinion that the maintenance of a stable general level of money wages is . . . the most advisable policy . . . There are advantages in some degree of flexibility in the wages of particular industries so as to expedite transfers . . . But the money-wage level as a whole should be maintained as stable as possible, at any rate in the short period. . . . In the long period . . . we are still left with the choice between a policy of allowing prices to fall slowly with the progress of technique and equipment whilst keeping wages stable, or of allowing wages to rise slowly whilst keeping prices stable. On the whole my preference is for the latter alternative. [pp. 270–71]

To Keynes, the price level was determined by the wage level—or more inclusively, the "cost-unit" a weighted average of the rewards of the factors entering into marginal prime cost—and the scale of output (p. 302). Changes in the money supply, or changes in the level of government spending exert their influence on the price level *through* their influence on the cost-unit, and the law of diminishing returns. If the Aggregate Production Function is as near linear as some maintain, (see R. G. Bodkin, E. Kuh, T. A. Wilson and Otto Eckstein), i.e., little diminishing returns or "bottleneck" effect, this does not vitiate Keynes' analysis. Rather it would mean that the level of Aggregate Demand, or of the money supply, can have no effect on the price level except through their effects upon the cost-unit, always excepting, as Keynes did, conditions of complete full employment where "true inflation" can occur (p. 303).[5] Perhaps we should avoid this extreme view, which negates the macroeconomic importance of so much that we teach in micro-theory. However, it is probably true that inflation control is much more a matter of preventing simultaneous upward shifts in Aggregate Supply and Demand, from upward shifts in the cost-unit, than it is a mechanistic matter of diminishing returns.

Eisner's analysis seeks to demonstrate that tax and interest rate hikes are weak anti-inflationary policies because they have only minor impacts on demand. He touches on the possibility that they are perverse only in a footnote: " . . . higher costs of money, like sales and excise taxes and, in the long run, business income taxes as well, tend to raise prices by raising costs of production" (p. 904). Might not the price increasing effects of higher interest and tax rates outweigh their price reducing effects, given the weaknesses Eisner highlights? Should not theorists consider this fully and econometricians sift the data in search of an answer? Economists seldom consider this possibility, however.[6] Here again we see the schizophrenia that afflicts economic theory. In the theory of the firm we discuss interest and taxes as not unimportant costs of production, and in the theory of tax incidence we depict the taxed businessman attempting to shift his taxes forward or backward, but we fail to develop the macroeconomic implications of all this.

---

[5] We approached, but at no time achieved such "full" employment in the Vietnam expansion. At all times there was a margin of unemployed resources and government and private demands expanded together. This weakens Eisner's argument regarding the impotence of monetary policy (p. 902). A few figures: in 1968 some 2,817,000 were unemployed out of a civilian labor force of 78,737,000, or 3.6 percent. Six hundred thousand more could have been employed before the unemployment rate would have equalled the 2.9 percent recorded in 1953, and 1,872,000 more would have had to be employed to equal the 1.2 percent rate of 1944 (see Council of Economic Advisors p. 252).

[6] Exceptions would include S. Weintraub (p. 149), P. Davidson, G. Brennan and D. A. L. Auld, whose analysis ". . . . demonstrates the possibility that the sales tax may be subject to cumulative shifting: first shifted forward to customers . . . then . . . back on firms in higher income claims to match the higher cost of living—then forward . . . and so on" (see Brennan and Auld, p. 525). A formal argument to support the conventional assumptions is seldom presented. However, see Earl Rolph regarding the price *decreasing* effect of higher sales taxes and G. Horwich on why interest hikes are non-inflationary. (See also my comment and Horwich's reply.) Richard Musgrave held that Rolph's conclusions are only justified in the classical model of quantity theory and pure competition, and that in the real world of imperfect competition, forward shifting will be the "typical case" (1953, p. 514). In his well-known text, Musgrave treats of the classical case (1959, pp. 364–71) but concludes that in the real world, " . . . A potential inflationary gap . . . may be closed by an increase in consumption taxes only while permitting some rise in the price level, reflecting the increase in cost due to tax" (1959, pp. 447–48). He holds, however, that this tax push inflation is " . . . a once-and-for-all increase, to be distinguished from the continuous increase that results if the inflationary gap is not closed," (p. 447) thus missing the possibilities highlighted by Brennan and Auld. Musgrave's conclusion that the Corporate profit tax is also fully shifted, (see Marion Krzyzaniak and Musgrave 1963) likewise has tax push implications, as his critic Robert J. Gordon points out (1967, p. 732). However, neither carry the point further in their subsequent exchange.

The theory that raising the tax and interest payments of business, and thereby causing some unemployment, reduces inflation involves the assumption of backward shifting to taxes and interest payments to factor suppliers. This is certainly possible and probably occurs to some extent over time, but that this is the sole, or even the major, impact of these policies upon incomes and prices appears highly dubious. Given the organization of producer interest groups and the failure of consumers to organize, forward shifting via higher prices appears the more likely outcome.

The theory that tax and interest rate hikes are inflationary in net effect is somewhat analogous to the balanced budget multiplier theory, but harder to quantify.[7] Increased government spending is very expansionary if it is financed by deficits and new money creation. It is still expansionary, but less so, if it is financed by higher taxes. Similarly, increased government expenditures, (in a nearly fully employed economy) are highly inflationary if they are financed by deficits and new money creation. Are they not still inflationary, but less so, if they are financed by higher taxes and the rate of interest is allowed to rise?

Rather than turn the world over to know-nothings, Keynesians should cross over the bridge Keynes built for them.

REFERENCES

P. S. Anderson, "Wages and The Guideposts: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 351–54.

R. G. Bodkin, "Real Wages and Cyclical Variations in Employment: A Re-Examination of the Evidence," *Can. J. Econ.*, Aug. 1969, *2*, 353–74.

G. Brennan and D. A. L. Auld, "The Tax Cut as an Anti-Inflationary Measure," *Econ. Rec.*, Dec. 1968, *44*, 520–25.

P. Davidson, "Rolph on the Aggregate Effects of a General Excise Tax," *Southern Econ. J.*, July 1960, *27*, 37–42.

R. Eisner, "Fiscal and Monetary Policy Reconsidered," *Amer. Econ. Rev.*, Dec. 1969, *59*, 897–905.

R. J. Gordon, "The Incidence of the Corporation Income Tax in U.S. Manufacturing, 1925–62," *Amer. Econ. Rev.*, Sept. 1967, *57*, 731–58.

———, "Incidence of the Corporation Income Tax: Reply," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1360–67.

G. Horwich, "Tight Money, Monetary Restraint, and the Price Level," *J. Finance*, Mar. 1966, *21*, 15–33.

———, "Tight Money as a Cause of Inflation: Reply," *J. Finance*, Mar. 1971, *24*, 156–57.

J. H. Hotson, "Neo-Orthodox Keynesianism and the 45° Heresy," *Nebr. J. Econ. Bus.*, autumn 1967, *6*, 34–49.

———, "Tight Money as a Cause of Inflation: Comment," *J. Finance*, Mar. 1971, *24*, 152–56.

J. M. Keynes, *The General Theory of Employment, Interest and Money*, London 1936.

M. Krzyzaniak and R. A. Musgrave, *The Shifting of The Corporation Income Tax*, Baltimore 1963.

——— and ———, "Incidence of the Corporation Income Tax: Comment," *Amer. Econ. Rev.*, Dec. 1968, *58*, 1358–60.

E. Kuh, "Unemployment, Production Functions, and Effective Demand," *J. Polit. Econ.*, 1966, *74*, 238–46.

R. A. Musgrave, "General Equilibrium Aspects of Incidence Theory," *Amer. Econ. Rev. Proc.*, May 1953, *43*, 504–17.

———, *The Theory of Public Finance*, New York 1959.

G. Perry, "Wages and the Guideposts," *Amer. Econ. Rev.*, Sept. 1967, *57*, 897–904.

———, "Wages and the Guideposts: Reply,"

---

[7] Much would seem to depend upon which tax is increased. Similarly, borrowers differ greatly in their ability to pass on their interest costs by increasing the price of whatever it is they sell. In a world of strong unions, however, it would seem that even the personal income tax is shiftable with a lag. For some discussion of these matters and attempts at quantification, see Hotson (1967).

Amer. Econ. Rev., June 1969, *59*, 365–70.
E. A. Rolph, *Theory of Fiscal Economics*, Berkeley 1954.
A. W. Throop, "Wages and the Guideposts: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 358–65.
M. L. Wachter, "Wages and the Guideposts: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 354–58.
S. Weintraub, *An Approach to the Theory of Income Distribution*, Philadelphia 1958.
T. A. Wilson and O. Eckstein, "Short-Run Productivity Behavior in U.S. Manufacturing," *Rev. Econ. Statist.*, Feb. 1964, *48*, 41–54.
Council of Economic Advisors, *Economic Report of the President*, Washington, Jan. 1969.

Fiscal and Monetary Policy Reconsidered: Comment

By BARBARA HENNEBERRY AND JAMES G. WITTE*

According to Robert Eisner in a recent issue of this *Review*, the keeping of a tight rein on the money supply can be expected to be an inadequate defense against inflation in the presence of a rise in government spending.[1] He then postulates a situation wherein government spending rises and taxes are not increased sufficiently to counter the expansionary effects on aggregate demand of this increment in government purchases. After further postulating that the nominal quantity of money is held constant, Eisner then points out that if the demand for real money is negatively related to the rate of interest, the velocity of circulation will increase. Under the assumed conditions of full employment, this will result in a rise in the price level. Thus, Eisner concludes, contemporary economic theory implies limitations on monetary policy (as well as tax policy) in combatting inflation.

Two points are worthy of note, and they both constitute important implications of the "modern" economic theory so dear to Eisner's heart. The first point is that as long as real commodity demand is a decreasing function of the rate of interest, there is some reduction in the quantity of money which would close the inflationary gap. Monetary policy is not logically constrained to zero or positive changes in the quantity of money. To use the terminology of J. R. Hicks and Alvin Hansen, the higher the (absolute) value of interest-rate elasticity of the *IS* function, the smaller need be such reduction, but there is some reduction which will offset rise in velocity induced by the increase in aggregate demand. Monetary policy, then, need not be rendered ineffective by the sheer existence of an interest-elastic *LM* function as Eisner seems to imply. After all, velocity is a function of the interest rate in any Keynesian model, but that does not imply that some reduction in the money supply could not serve to offset an interest-induced rise in velocity.

Second, a given shift in aggregate demand, with a constant supply of money, will produce a *one-time* increase in velocity and prices. Yet, as Eisner correctly points out, the problem during the 1967–69 period was that of a continuing rise in prices which has lasted well beyond the gradual elimination of the fiscal stimulant. Thus, in order to explain the continuing rise in prices one must account for a further exogenous disturbance of a more continuing nature. We submit that the continuing disturbance could be the employment of a monetary policy which permitted the money supply to rise steadily at quarterly rates which tend to exceed the growth in real *GNP*.[2] While the effect of the fiscal stimulant is to produce a new, higher equilibrium value of the rate of interest which closes the inflationary gap, the attainment of such an equilibrium rate of interest can be prevented by expansionary behavior on the part of the central bank. Central bank action which inhibits the tendency of the rate of interest to rise to its inflationary-gap-closing levels results in a positive rate of inflation that continues at least as long as the growth in the money supply continues, and this, we submit, is what has constituted the phenomenon exhibited over the time period in question. Contrary to Eisner, the inflationary problem results, not from the *inadequacy* of monetary policy but from the *improper use* of monetary policy.

---

* Former assistant professor of economics, Ball State University and professor of economics, Indiana University, respectively.

[1] The authors have chosen to comment only upon Eisner's monetary statement; they are in substantial agreement with him on his fiscal policy position.

[2] The referee has indicated to us his belief that our test of the direction of monetary policy, viz. the rate of growth of the money supply relative to the growth of real *GNP* may be inappropriate. He suggests, instead, that the relevant test of the tightness of a monetary policy in an inflationary situation "is not whether the money supply increases more rapidly than the growth in real *GNP* but whether it increases more rapidly than the growth in the money value of *GNP*"; by this criterion, he says, "monetary policy was in the proper direction but unsuccessful." We must disagree with this definition. The ratio of nominal *GNP* to the money supply is equal to the velocity of circulation. Use of such a criterion would mean that a tight money policy is in effect *whenever* velocity rises. We would then have to characterize monetary policy in the 1946–51 period as tight, despite the fact that the Treasury Bill rate was pegged by continuous open-market operations. By our definition, monetary policy during the early postwar period was clearly one of excessive ease; it is our contention that, in 1946–51 as in 1966–68, excessively easy monetary policy led to substantial increases in the general price level.

REFERENCE

R. Eisner, "Fiscal and Monetary Policy Reconsidered," *Amer. Econ. Rev.*, Dec. 1969, *59*, 897–905.

Fiscal and Monetary Policy Reconsidered: Comment

By KEITH M. CARLSON*

Robert Eisner has recently entered the debate on the relative potency of monetary and fiscal actions. He demonstrates the ineffectiveness of the 1968 tax surcharge in checking inflation, then goes on to assert that tight money would be similarly ineffective. This paper considers Eisner's analysis as it pertains to the inadequacy of monetary policy. First, it is shown that his conclusions do not necessarily follow from his own model. Second, using parameter estimates representative of other studies, it is demonstrated that Eisner's conclusions are not substantiated by the empirical evidence.

I. Solving Eisner's Model

Eisner summarizes his model of aggregate demand as follows:

Private Commodity Demand

$$C = C(Y, Y^e, i, A, M) \tag{1}$$

Real Money Demand

$$M^D = M^D(X, i, A) \tag{2}$$

Output Equilibrium

$$X = C + G \tag{3}$$

Money Market Equilibrium

$$M = M^D \tag{4}$$

Money-Price Identity

$$M = \frac{Q}{P} \tag{5}$$

The symbols are defined as:

C = real consumption plus real investment demand
Y = current real income after taxes
Y^e = expected future real income after taxes
i = real rate of interest[1]
A = net real value of nonmonetary assets held by private sector
M = real cash balances held by private sector
M^D = real cash balances demanded by private sector
X = real output
G = real government demand for goods and services
Q = nominal quantity of money
P = commodity price index

This is a system of five equations with C, i, M, M^D, and P endogenous; Y, Y^e, X, G, Q are considered exogenous. To formally close the model (which Eisner does not do) another equation is added:

Nonmonetary Asset Identity

$$A = \frac{1}{iP}a \tag{6}$$

In this equation, a is exogenous and represents the number of nonmonetary assets, each of which can be considered as a perpetuity paying one dollar per year. Eisner places considerable emphasis on the effects on A of changes in i and P without explicitly specifying a relation like (6).

Differentiating equations (1)–(6) totally yields the following:

$$dC = C_Y dY + C_{Y^e} dY^e + C_i di + C_A dA + C_M dM \tag{7}$$

$$dM^D = M_X dX + M_i di + M_A dA \tag{8}$$

$$dX = dC + dG \tag{9}$$

$$dM = dM^D \tag{10}$$

$$dM = \frac{1}{P}dQ - \frac{Q}{P^2}dP \tag{11}$$

---

* Federal Reserve Bank of St. Louis. The views presented here do not purport to represent those of the Board of Governors of the Federal Reserve System or the Federal Reserve Bank of St. Louis.

[1] Eisner does not identify his interest rate as nominal or real. Since price expectations are not mentioned in the article, he apparently assumes that the nominal and the real rate are identical.

$$(12)\qquad dA = \frac{1}{iP}da - \frac{a}{i^2P}di - \frac{a}{iP^2}dP$$

To solve the system, dY, dY^e, and dX are all set equal to zero in accordance with the full employment assumption. Two cases are considered with respect to the method of financing an increase in government expenditure. The first is bond financing, or $dG=1/iP\ da$ (thus $dQ=0$); the second is money financing, or $dG=1/P\ dQ$ (thus $da=0$).[2]

Case 1: Bond Financing

To show the effects of bond financing on the price level, the system is reduced to the following two equations:

Commodity Equilibrium

$$(13)\qquad \left(\overset{(+)}{C_A}\frac{a}{i^2P} - \overset{(-)}{C_i}\right)di = -\left(\overset{(+)}{C_A}\frac{a}{iP^2} + \overset{(+)}{C_M}\frac{Q}{P^2}\right)dP + \overset{(+)}{(1+C_A)}dG$$

Money Equilibrium

$$(14)\qquad \left(\overset{(+)}{M_A}\frac{a}{i^2P} - \overset{(-)}{M_i}\right)di = \left(\overset{(+)}{\frac{Q}{P^2}} - \overset{(+)}{M_A}\frac{a}{iP^2}\right)dP + \overset{(+)}{M_AdG}$$

The signs of the partial derivatives (or terms including levels of variables), as postulated by Eisner, are written above those derivatives. Clearly, the commodity equilibrium equation, when charted in the (i, P) plane (following Don Patinkin), is negatively sloped. The money equilibrium curve may be either positively or negatively sloped. Conventionally, it is assumed to be positively sloped, but if $M_A(a/iP^2) > Q/P^2$ (i.e., a strong real nonmonetary asset effect on real money demand), the money equilibrium curve is negatively sloped.

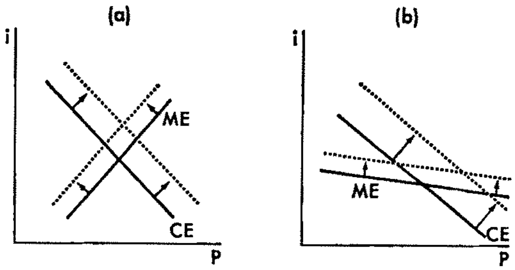


FIGURE 1. CASE 1: BOND FINANCING OF GOVERNMENT SPENDING

The effect of a change in government spending financed by bond issue is shown in the two panels of Figure 1. Panel (a) shows the effect of an increase in government spending for the positively sloped *ME* subcase. For an increase in government spending financed by bond issue, the effect on i is unambiguously positive, while the effect on P is ambiguous, depending on the relative magnitude of shifts of the two curves. Panel (b), with a negatively sloped *ME* curve, shows that the effect of an increase in government spending is ambiguous for both i and P.

As a result of solving the model, there is no basis, on strictly logical grounds, for accepting Eisner's conclusion that an increase in government spending, with the nominal money stock held constant, leads to an increase in the price level.[3] Any such conclusion requires consideration of the empirical evidence.

Case 2: Money Financing

Since Eisner goes on to argue that monetary policy would be ineffective in combatting inflation, a second case is considered before examining the evidence. Rather than examine the effect of a decline in the money stock, the effect of an increase in the money

---

[2] These two cases indicate that only two of the three policy variables, dG, da, and dQ, can be determined independently; thus the third becomes endogenous in the model. The importance of this government budget restraint has been emphasized by Carl Christ.

[3] Strangely enough, Eisner shows that he is aware of these possibilities. But the discussion of these possibilities is relegated to a footnote.

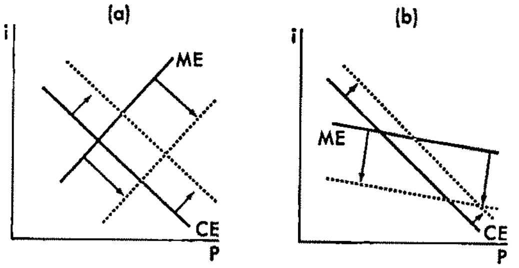


FIGURE 2. CASE 2: MONEY FINANCING OF GOVERNMENT SPENDING

stock on the price level is analyzed instead. Case 2 involves an increase in government spending accompanied by monetary expansion.

The system (7)–(12) is solved as before, except that $dG = 1/P\, dQ$ and $da = 0$. This case reduces to the following:

Commodity Equilibrium

$$\left(\overset{(+)}{C_A \frac{a}{i^2P}} - \overset{(-)}{C_i}\right) di = -\left(\overset{(+)}{C_A \frac{a}{iP^2}} + \overset{(+)}{C_M \frac{Q}{P^2}}\right) dP + \overset{(+)}{(1 + C_M)} dG \tag{15}$$

Money Equilibrium

$$\left(\overset{(+)}{M_A \frac{a}{i^2P}} - \overset{(-)}{M_i}\right) di = \left(\overset{(+)}{\frac{Q}{P^2}} - \overset{(+)}{M_A \frac{a}{iP^2}}\right) dP - dG \tag{16}$$

Equations (15) and (16) are the same as (13) and (14) except for the coefficients on dG. The effects of a change in government spending financed by money issue are shown in Figure 2.

The subcase shown in panel (a) demonstrates that an increase in government spending accompanied by monetary expansion is unambiguously positive in its effect on P. The effect on the interest rate now becomes ambiguous. Likewise for the negatively sloped ME subcase, government spending financed by money issue yields an unambiguously positive effect on P. And, for this subcase, the effect on interest rates is unambiguously negative.

After examining four subcases reflecting different sets of assumptions, the conclusion follows that the effect of government spending on prices depends critically on whether or not it is accompanied by monetary expansion. If the money stock is held constant and government spending is increased, there *may* be an increase in prices, but it is not until the assumption of unchanged money is relaxed that the effect becomes unambiguous.

II. Some Empirical Evidence

Eisner's conclusions about the factors underlying inflation do not hold even within the framework of his own model. Nevertheless, it is of interest to examine the magnitudes involved. To do so, parameter estimates from a recent econometric study are used (see the paper by Carlson and Denis Karnosky).[4]

The values assigned to the parameters and relevant variables are as follows:

Impact	Steady State	Levels
$C_i = -2.55$	$C_i = -7.07$	$a/iP = 1020.00$
$C_A = .13$	$C_A = .40$	$i = 3.87$
$C_M = .87$	$C_M = 2.67$	$P = 1.18$
$M_i = -1.96$	$M_i = -20.40$	$Q/P = 151.00$
$M_A = .02$	$M_A = .21$	

The levels of variables are for late 1967. The parameter estimates are based on regressions for the 1953–67 period. Both impact and steady state coefficients are shown.

Inserting these values into the relevant equations for the two cases, the following results are obtained:[5]

---

[4] The study from which these estimates were taken was based on a Keynesian income-expenditure model. Parameter estimates conform with those of other studies (see Christ).

[5] The magnitudes of the coefficients were such that the stability conditions were satisfied in all cases.

	Impact Effect	Steady State Effect
Case 1:		
$dP/dG\|_{dQ=0}$	.0013	.0018
$di/dG\|_{dQ=0}$	.0228	.0014
Case 2:		
$dP/dG\|_{da=0}$	.0088	.0085
$di/dG\|_{da=0}$	−.0027	−.0193

Because the magnitudes of the multipliers are small, the difference in effect on prices is not immediately evident. However, the bond financing case shows a .11 percent impact effect on prices, while the money financing case represents a .75 percent increase (where both percents are calculated from a late 1967 base). The price effects are of the same relative magnitude for the two cases in the steady state.

These results can be applied in a similar way so as to shed light on Eisner's conclusion. From fourth quarter 1965 to fourth quarter 1969, real government purchases of goods and services increased $15.5 billion. Treating the increase for that period as a once and for all increase of half that amount in late 1965, and applying the steady-state multiplier for the bond financing case to that figure of $7.8 billion, yields an increase in the price level of .014, or 1.26 percent, from late 1965.[6] The actual increase was .19, or 17.04 percent. Though apparently correct in direction, Eisner's strong conclusion about the inflationary impact of government spending (when financed by bond issue) receives little support from the evidence.

III. Conclusion

Eisner concluded that tax policy and monetary policy were ineffective as tools for fighting inflation. His conclusions on tax policy may be valid; this paper focuses on monetary policy, with special emphasis on its association with financing of government expenditures. Using Eisner's model, it was found that his conclusions do not necessarily follow. Since any definite conclusions ultimately depend on the values of the partial derivatives, a set of parameter estimates was used to calculate the effects on the price level of an increase in government spending financed by bond issue and by money issue. These calculations, though admittedly approximate, indicated that government spending has a much smaller effect on the price level if unaccompanied by monetary expansion. Eisner appears to be correct in the direction of the effect of deficit-financed expenditures on the price level, but empirical evidence indicates the magnitude of the effect is very small.

REFERENCES

K. Carlson and D. Karnosky, "The Influence of Fiscal and Monetary Actions On Aggregate Demand," working paper no. 4, Federal Reserve Bank of St. Louis, Mar. 1969.

C. Christ, "A Short-Run Aggregate Demand Model of the Interdependence and Effects of Monetary and Fiscal Policies with Keynesian and Classical Interest Elasticities," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 434–43.

R. Eisner, "Fiscal and Monetary Policy Reconsidered," *Amer. Econ. Rev.*, Dec. 1969, *59*, 897–905.

D. Patinkin, *Money, Interest, and Prices*, 2d ed., New York 1965.

---

[6] This procedure is approximate, but the contribution of deficit-financed increases in government expenditures to the explanation of changes in the price level seems small enough (less than 10 percent) to justify rejection of Eisner's main conclusion.

Fiscal and Monetary Policy Reconsidered: Reply

By ROBERT EISNER*

The four comments fall into two groups. Bent Hansen and John Hotson offer little or no objection to my critique of monetary policy. As to recent and conventional fiscal policy, they also apparently share my reservations but have some strictures of their own to add. Keith Carlson, James Witte and Barbara Henneberry seem content with the rebuke to fiscalists but seek to raise very considerable objection to corresponding questioning of monetary policy. One may infer that, for perhaps varied reasons, all contributors are pleased to have fiscal policy raked over the coals, but affront to monetary gods is offensive to some of them.

I have little quarrel with Hansen's arguments and have indeed developed several of his points elsewhere (1971). I do, as he anticipates, rather doubt a substantial negative elasticity of the supply of labor with respect to real after-tax income, but will await with interest the results of any new empirical studies of this subject. I believe there is considerable substance to his suggestion that supply functions for both labor services and private products are such that higher taxes may have serious cost-inflationary effects and that certain kinds of government expenditures may be cost-deflationary. And I am similarly appreciative of Hotson's arguments that tax and interest rate increases are cost-inflationary, but more skeptical of the wisdom and efficacy of "guidepost" restraint on wages.

With my monetarist critics I have more argument. First, Witte and Henneberry are of course correct in arguing that a demonstration that holding the quantity of money constant may not be sufficient to prevent price inflation "does not imply that some reduction of the money supply could not serve to offset an interest-induced rise in velocity." It does not logically follow, however, that "as long as real commodity demand is a decreasing function of the rate of interest, there is some reduction in the quantity of money which would close the inflationary gap." This, after all, depends upon how rapidly a decreasing function we have and just how large an inflationary gap must be closed. One may even conceive a situation in which the government insists upon keeping real commodity demand at an inflationary level while reducing the quantity of "money," however it was originally defined, to zero. The consequence might then be continued inflation, with non-banking institutions furnishing debt instruments which could serve as substitutes for the money which the Witte and Henneberry-dominated Fed might have taken out of circulation.

I can only agree with the anonymous referee who points out that the test of tightness of monetary policy must involve the relation of the quantity of money to the money value rather than the real value of *GNP*. Witte and Henneberry, along with a surprising number of monetarists, strangely seem to ignore that the nominal quantity of money is most essentially an endogenous element of our monetary system as well as most if not all others. Money is allowed to vary largely with "the needs of trade," and will certainly have a very strong tendency to rise whenever the money value of *GNP* rises. Unless the quantity of money increases more than the money value of *GNP*, Witte and Henneberry (and others) would clearly have their cause and effect reversed. This partial bending of the monetary authority to the needs of an inflationary economy, the product of repeated lessons of the consequences of short-run liquidity crises, may be considered as one of the elements in a lag process that could have been expected to contribute to "a continuing rise in prices which has lasted well beyond the gradual elimination of the fiscal stimulant."

---

* Northwestern University and National Bureau of Economic Research. The author is grateful for comments by Robert W. Clower on a draft of this reply.

We may note that Witte and Henneberry state that "they are in substantial agreement with [Eisner] on his fiscal policy position." Carlson's similar, if not quite definitive, acquiescence in my fiscal critique—[Eisner's] "conclusions on tax policy may be valid"—puts him in a position he may well wish to reconsider. For Carlson espouses an argument, which has recently gained considerable currency if far from universal acceptance, that there is one particular kind of paper, variously defined as legal tender, properly signed and valid checks and, in some instances, properly valid pages of savings account books, which is all-decisive. Other pieces of paper, no matter how much they look like the magic ones which can do all, have little or nothing to do with inflation. Carlson's analysis of the relative effects of bond and money-financed deficit spending along with his concession to my critique of fiscal policy bring some of these views into a clearer, perhaps glaring light.

For what Carlson is saying comes down to the following. If the government increases expenditures and raises money for these expenditures by taxes for which the private sector gets nothing in return other than pieces of paper called tax receipts, aggregate demand (and at least under conditions of full employment, prices) will rise. If the same increase in government expenditures is accompanied not by increases in pieces of paper marked tax receipts but by pieces of paper called money, Carlson would look for a much more inflationary impact—and here most of us would concur. But if the increase in government expenditures involves leaving the public not with more tax receipts and not with more money but with more government bonds, Carlson develops grave doubts as to their inflationary impact. One can indeed concede that having a thousand dollar bill from the government will be more conducive to private demand than having a thousand dollar tax receipt. But is a thousand dollar bond so much worse than a thousand dollar bill as to be worse than a thousand dollar tax receipt? Carlson concedes that increased government expenditures, which leave the private sector with no more net assets denominated in monetary terms, can be inflationary. How then can he deny that similar expenditures, which leave the private sector with monetary assets that are deferred obligations of the government rather than an instant obligation, will somehow be deflationary? Indeed, since neither Carlson nor I specify the duration of the government debt issues used to finance a budget deficit, Carlson is left in the strange position of arguing that a thousand dollar noninterest-bearing bill, which he would call money, is very inflationary but a thousand dollar interest-bearing bill which may differ from the former only in that it will not become money until tomorrow, and offers a premium of sixteen cents for the delay, may be deflationary!

Carlson insists (and Hansen also notes) that my model does not absolutely preclude a decline in prices with bond-financed deficit spending. What I did was to indicate the extreme behavioral assumption necessary for this result. (" . . . it would be to argue that the increase in interest rates necessary to reduce commodity demand despite the increase in the net value of assets would not be sufficient to reduce the real demand for money, implying a shift from commodities to money as interest rates and bond holdings rise" p. 903.) Carlson's further formalization of my model along with his own rough empirical estimates may indeed be used to substantiate the argument I advanced. Carlson points out, setting up commodity equilibrium and money equilibrium curves, that effects on price are "ambiguous, depending on the relative magnitude of shifts of the two curves." What this comes down to in the case of bond financing is the question of whether, setting $dP=0$, di/dG in Carlson's equation (14) is greater than di/dG in equation (13). This is to say that

$$\frac{dP}{dG} \gtrless 0 \quad \text{as}$$

$$\frac{1+C_A}{C_A\dfrac{a}{i^2P}-C_i} \gtrless \frac{M_A}{M_A\dfrac{a}{i^2P}-M_i}$$

Thus, as indicated qualitatively in my original paper, for deflationary consequences of debt-financed government spending, M_A must be so large and the absolute value of M_i so small as to bring about "a shift from commodities to money as interest rates and bond holdings rise."

Carlson's own empirical estimates suggest how far from this strange condition we are. For substituting the parametric estimates he furnishes we have for the critical inequality:

$$\frac{1.13}{.13\left(\frac{1020}{3.87}\right)+2.55} \overset{?}{>} \frac{.02}{.02\left(\frac{1020}{3.87}\right)+1.96}$$

or .3007 > .0028 for the "impact" result, and

$$\frac{1.40}{.40\left(\frac{1020}{3.87}\right)+7.07} \overset{?}{>} \frac{.21}{.21\left(\frac{1020}{3.87}\right)+20.40}$$

or .0124 > .0028 for the steady state, indicating that in either case bond-financed increases in government expenditures are inflationary.

Indeed Carlson's estimates suggest that if the demand for money were perfectly inelastic, that is, $M_i=0$, the effect of increased government debt in the hands of the public would still not raise the demand for money sufficiently to lower prices. Aside from any other measurement or statistical problems in the unpublished manuscript which Carlson cites for his estimates, I might point out that his a/iP is in fact the money value of total private wealth. The elasticity of demand for money with respect to total wealth is likely to yield a higher figure for M_A than the correctly estimated derivative of the demand for money with respect to holdings of government bonds, a closer substitute for money. One may conjecture that I conceded too much in even allowing $M_A>0$; M_A, at least with respect to short-term bills, may well be negative. One can hardly entertain seriously the implication of Carlson's position, contrary to the empirical results he cites, that M_A is both positive and so large as to make bond-financed deficit-expenditures deflationary when similar tax-financed expenditures are inflationary.

Another way of exposing the remarkable role ascribed by Carlson (and others) to money as opposed to other financial assets is to compare the critical last terms of his money equilibrium equations (14) and (16). For in (16), the money-financing case, there must be an adjustment of interest rates and prices to compensate for $-dG$, the reduction in the excess demand for money brought on by its increased supply. In (14), the bond-financing case, there must be an adjustment to a presumed increase in excess demand for money equal to $M_A dG$. One may ask indeed how similar bonds must become to money before M_A in (14) approaches the value of -1 in (16), and C_A approaches C_M, in the last terms of (13) and (15). Suppose those one thousand dollar bills alluded to earlier were debt instruments maturing five minutes after they were received? Sophisticated monetarists may try to deny this, but their critical implicit notion in making money sovereign is that it is somehow unique, *both* as a means of immediate exchange and in all the other qualities attributed to financial assets. Hence, adding other financial assets can never free money for more of its presumably unique role in transactions, and there can be no substitution of other financial assets which would permit the same amount of money to finance an increased dollar volume of transactions.

It is difficult to respond to Carlson's admittedly "approximate" procedure for estimating the amount of inflation that might have occurred had the quantity of money been kept constant from 1965 to 1969, since the underlying model and basis for his estimates are not presented in his paper. It may be pointed out, however, that the figures furnished by Carlson suggest that even if increased government expenditures had been financed entirely by the creation of money, presumably his case 2, the actual increase in prices from 1965 to 1969 would have been only .0092/.0019 times the 1.35 percent increase which he argues would have taken place if the quantity of money had been kept constant (his case 1). Thus Carlson's numbers suggest that even the most inflationary financing of increased government expenditures would have accounted for an inflation of only about 7 percent, as against the 17 percent acknowledged by Carlson as the increase over this period. Since financing larger government expenditures entirely by an increase in the quantity of money would actually have been more inflationary, by both Carlson's reasoning and mine, than what actually occurred, there is clearly something lacking or wrong in Carlson's model or numbers.

I return to my original thesis. The countercyclical fiscal policy advanced during the 1965–69 period, let alone that actually adopted, should not have been expected to be successful against the war-induced inflation. And just as surely no reasonable monetary policy, even including the perhaps unreasonable policy of trying to keep the nominal quantity of money constant, would have been successful either.

REFERENCE

R. Eisner, "What Went Wrong?", *J. Polit. Econ.*, May/June, 1971, *79,* 629–41.

More on an Empirical Definition of Money: Note

By DAVID T. HULETT*

This study was conducted to evaluate George Kaufman's extension of the Friedman and Meiselman technique for an empirical definition of money. This method defines as money that financial aggregate which satisfies two criteria: 1) it exhibits the highest correlation with *GNP*, and 2) the correlations between *GNP* and each of the components considered separately do not exceed that between *GNP* and the aggregate. The components are thus substitutes—the public alters the composition of its portfolio due to changes in supply conditions, while keeping its portfolio size constant relative to *GNP*. (See Friedman and Schwartz, ch. 2; and J. R. Hicks, p. 49.) The set of assets heretofore considered include liquid financial assets. Friedman and Meiselman discovered that the dual criteria were best satisfied by the sum of currency and all privately held deposits at commercial banks (pp. 182–84).

Kaufman examined the proposition that if money is a factor in determining *GNP*, its effect may be delayed by as long as a year. From correlations between *GNP* and various financial aggregates which led *GNP* by +4 to −2 quarters, he found that the best definition of money depends on the number of quarters by which the financial measure leads or lags *GNP*. In general, the broader aggregates perform better when they are observed two or more quarters before income while the narrow definition performs best when observed concurrently with income (see Kaufman, pp. 86–87, Tables 1 and 2).

Kaufman examines the impact of changes in the monetary aggregate on changes in income in a particular current or future quarter, a procedure which is appropriate if money's effect on income occurs with a discrete time lag.

The present study extends the Kaufman analysis, allowing the effect of money on *GNP* to be distributed over several quarters by examining regressions of the following form:

$$GNP_t = a_0 + a_1M_t + a_2M_{t-1} + a_3M_{t-2} + a_4M_{t-3}$$

The distributed lag model examines the contention that the effect of changes in money spread slowly through the economy and that the long-run impact on *GNP* may build up over several quarters. Put differently, this approach investigates the proposition that knowledge of the time path of monetary changes might be more useful than a single quarter's change in determining the change in *GNP*.

I. The Correlation Tables

Data were collected in accordance with Kaufman's technique. Four end-of-month observations of various financial stocks were averaged to arrive at an average quarterly stock. Financial aggregates were then constructed : $M1$ = currency plus demand deposits; $M2 = M1$ plus time deposits at commercial banks; $M2.5 = M2$ plus deposits at mutual savings banks; $M3 = M2.5$ plus savings and loan shares; $M4 = M3$ plus savings bonds and postal savings deposits; and $M5 = M4$ plus private holdings of *U.S.* government short-term marketable securities. (A second method added currency last, following Kaufman's observation that currency is more closely related to income observed in earlier periods (p. 86), implying that income determines the demand for currency but currency is not a driving force determining income.) The financial variables and *GNP* were transformed to logarithms and expressed as first differences. The data here extend Kaufman's observations by two recent years; they cover the period 1953–1968. The correlations

---

* Economist, Office of Management and Budget. This note has benefitted from discussions with J. L. Pierce and B. Friedman. The empirical calculations were largely performed by J. Walton. The author is solely responsible for the views expressed.

TABLE 1—AGGREGATE CORRELATION MATRICES

	Quarters Money Leads *GNP*				
	+4	+3	+2	+1	0
1953–68					
Currency	−.01	.04	.14	.26	.40
$M1=\text{Currency}+DD$	.04	.26	.41	.51*	.49*
$M2=M1-TD$	.28	.41*	.49*	.47	.32
$M2.5=M2+MSB$	.29	.41*	.49*	.47	.31
$M3=M2.5+SC$	.28	.41*	.48*	.45	.27
$M4=MS+SB$	.31	.44*	.50*	.46	.30
$M5=M4+$ *U.S.*	.03	.07	.19	.30	.39
1953–59					
Currency	−.61	−.59	−.30	−.14	.32
$M1$	−.40	−.15	.24	.63	.70*
$M2$	−.17	.03	.35*	.55	.30
$M2.5$	−.17	.04	.37*	.55	.29
$M3$	−.15	.07	.41*	.60	.33
$M4$	−.15	.10	.46*	.64	.36
$M5$	−.37	−.65	−.44	−.10	.30
1960–68					
Currency	.29	.32	.30	.39	.43
$M1$	.34*	.61*	.59	.47*	.32
$M2$	.51*	.66*	.62*	.44	.21
$M2.5$	.53	.67*	.62*	.44	.20
$M3$	.50	.66*	.57	.35	.07
$M4$	.54	.67*	.59	.37	.10
$M5$	.10	.48	.67*	.61*	.40

Source: Federal Reserve Board.

The variables are quarterly first differences of natural logarithms. MSB = deposit at mutual savings banks, SC = shares at savings and loan associations, SB = savings bonds and postal savings deposits, and *U.S.* = private nonbank holdings of *U.S.* government marketable securities maturing within one year.

* The entries marked with an asterisk satisfy the criterion that the aggregate correlation coefficient exceeds those of *GNP* with its components separately.

and regressions have been calculated for the whole period and the two subperiods, 1953–1959 and 1960–1968.

Correlation matrices shown in Table 1 were calculated as in Kaufman's study on the assumption of a point-input point-output relationship. Only the aggregate correlations are shown, but the asterisk indicates that the second criterion was satisfied. The results are essentially those found by Kaufman.

First: For the entire period the narrow definition of money ($M1$) performs best and satisfies the dual criteria when observed concurrently or one quarter before *GNP*; $M4$, which includes all of the "liquid assets" except marketable short-term government bonds, performs best when observed two or three quarters before *GNP*. In fact, the short-term government bonds, and $M5$ which includes them, do very poorly by these tests. (Adding currency last raises the correlations by about .02. This table is not shown.)

Second: The results for the two subperiods differ substantially from each other. The large current effect of $M1$ appears only in the earlier subperiod while the effect of money two or more quarters ahead of income is derived largely from the recent subperiod. The correlations with *U.S.* government bonds and $M5$ are strongly positive only in the recent period as are those of currency, while the correlations with $M1$ do not dominate those of the broader definitions in any lead quarter of that subperiod.

TABLE 2—*F*-STATISTICS TO TEST THE EQUALITY BETWEEN COEFFICIENTS OF MONEY IN REGRESSIONS OF THE FORM $GNP = a + bM_{t-i}$ (CHOW TEST) FOR THE PERIODS 1953–59 AND 1960–68

	t+4	t+3	t+2	t+1	t
Currency	14.55	14.27	4.39	1.78	.77
*M*1	7.36	4.11	1.05	6.21	8.72
*M*2	2.63	.66	.61	4.16	1.16
*M*2.5	2.61	.66	.88	4.15	1.12
*M*3	2.40	.59	1.50	5.94	2.18
*M*4	2.12	.19	2.21	7.65	2.18
*M*5	5.44	19.77	11.97	3.05	.52

The *F*-statistic with (3, 59) degrees of freedom is significant at 3.15 (5%) and 4.98 (1%).

II. Stability of the Relationships

To test Kaufman's conclusion that "These findings do not differ greatly for the two subperiods" (p. 84), with the extended data, the Chow test (see J. Johnston, pp. 136–38) was performed on simple regressions of *GNP* on money for the three time spans. The resulting *F*-statistics are presented in Table 2. Clearly, many of the coefficients have changed significantly. The improved performance of currency and *M*5 are the most marked, while the decline in impact of *M*1 in periods t and t+1 and its improvement in t+3 and t+4 are also highly significant. When the subperiods are compared to the correlations for the whole period it is clear that the large immediate impact of *M*1 is due mostly to the period before 1960 and that the performance of the broader aggregates is perhaps more relevant today. An examination of the recent subperiod casts doubt on the timeliness of Kaufman's contention that *M*1 satisfies the dual criteria best (p. 86).

A developing financial sophistication of the household and business sectors over the recent years may have led to some of the differences between the two subperiods. For instance, there are over twice as many asterisks in the recent subperiod than in the earlier one, indicating that asset substitution is becoming more widespread. The increasingly competitive nature of the financial markets, particularly in the 1960's when the demand for credit drove short-term interest rates to historic levels, has alerted the public to the desirability of allocating assets according to conditions of supply (relative interest rates).

One explanation for the dramatic change in the role of short-term government securities since 1960 is that corporate holdings of bills for tax anticipation purposes have been brought more into line with the timing of income earned.[1] Prior to 1960, tax payments lagged behind the realization of profits by as much as a year, and the bills which were accumulated for this purpose lagged the economic activity (*GNP*) which gave rise to the profits. Some support for this proposition is given by the large positive correlations in the 1953–1959 period between *GNP* and short-term bonds observed one and two quarters later (see Kaufman, p. 80). Since 1960, tax payments have been made gradually more synchronous with profits earned, and the bills which are held as tax hedges have tended to be accumulated more nearly when the level of business activity is highest. Since tax transactions are one expense of doing business, this source of demand for liquidity properly belongs here.

In addition, corporate treasurers have used an increasing portion of their holdings of bills in their general liquidity portfolio. The rapid development of the negotiable time certificate of deposit (*CD*'s) since 1961 has contributed to this development. *CD*s are well suited to perform the tax hedge role: a business can solidify a customer relationship while earning interest on money which would not otherwise be free for other uses. Banks accommodate this need by bunching the *CD* maturities around tax dates. The behavior of the secondary *CD* interest rate attests to the similarity of *CD*s and other short-term marketable bonds.

---

[1] This idea was raised in discussions with B. Friedman.

$$\begin{aligned} RCD = {} & \underset{(6.53)}{.574\, RTB} + \underset{(5.01)}{.432\, RCP} \\ & + \underset{(5.75)}{35.976\,(CD/GNP)} - \underset{(-5.78)}{.590} \end{aligned}$$

$R^2 = .994$ $D.W. = 1.51$ Period: 1961–IV to 1968–IV $S.E. = .07$ $D.F. = 15$

In this quarterly regression, a *CD* demand function has been normalized on the secondary *CD* rate (*RCD*), and includes strong effects of the bill rate (*RTB*) and the commercial paper rate (*RCP*).[2] In this environment, including strong rate substitution and the increased availability of other assets which can perform the tax hedge role, bills are largely freed to perform the liquidity role, contributing to their increases in "moneyness" since 1960.

Further evidence which supports these tentative conclusions can be found in W. H. Locke Anderson's results. In his time period, 1948–60, corporate short-term bond holdings bear very little relationship to sales, his transactions variable, but are highly correlated with accrued tax liabilities as the present hypothesis implies. In some of his capital expenditure equations, the coefficients on short-term bonds and tax accruals are essentially equal in value but opposite in sign, indicating that only the excess of bonds over tax accruals is relevant for the real investment decision (see pp. 45–51, 75–76, especially equation (5)), another implication of the hypothesis suggested above. It would be interesting to examine these relationships for the period since 1960: the expectation is that they would no longer hold.

III. Distributed Lags

In an attempt to investigate further the implication of a lagged response of *GNP* to monetary aggregates and to try to achieve a more reliable prediction of *GNP*, distributed lags (rather than discrete lags) were estimated by regressing *GNP* against current and several prior observations on the monetary variables. In this framework, the total effect of money on *GNP* could be accumulated over several periods. The results of this experiment are displayed in Table 3.

The application of distributed lags does not change the conclusions drawn from the discrete lag experiment.

First: For the whole period and early subperiod, the narrow definition of money ($M1$) performs the best. Adding lags beyond $L=1$ to $M1$ does not improve its predictive power. In the 1953–59 period, the coefficient of determination is rather large for $M1$, but the application of distributed lags does not significantly improve on the correlations of Table 1. The sums of coefficients on lagged values of $M5$, which includes short-term securities, are sometimes negative, contrary to expectation.

Second: For the recent subperiod, the broader measures perform best. $M5$ has perhaps the best record: its estimation of *GNP* is about the best with only the current and two lagged observations. The large predictive power of $M1$ exhibited in the early subperiod has been eroded. These results essentially parallel those of the simple correlation table.

IV. The Role of an Empirical Definition of Money

The usefulness of the debate over the definition of money can be questioned on methodological grounds as well as on its relevancy to policy questions. The empirical method used to define money should correspond to the form of the money demand function in the underlying model. Leaving aside for the moment the well-known debate concerning the presence of absence of interest rates in this function (see Galper, Hamburger (1969), and Lee), there remains the issue of

---

[2] The observations in which *RCD* exceeded the Regulation Q ceiling have been omitted since the change in outstandings largely represents a runoff from the previous maturities in those periods. This regression is due to H. T. Farr.

TABLE 3—VALUES OF ADJUSTED R^2 FOR THE DISTRIBUTED LAG EQUATIONS

$$\Delta\, Ln\, GNP = a + \sum_{i=0}^{L} b_i \Delta\, Ln\, M_{t-i}$$

	$L=4$	$L=3$	$L=2$	$L=1$	$L=0$
1953–68					
Currency	.21	.19	.14	.15	.15
$M1=$ Currency$+DD$	.29	.28	.29	.29	.23*
$M2=M1+TD$	.21	.22	.22	.20	.09
$M2.5=M2+MSB$	.20	.21	.21	.20	.08
$M3=M2.5+SC$	.20	.21	.21	.18	.06
$M4=M3+SB$	.21	.23	.23	.20	.07
1953–59					
Currency	.38	.38	.21	.14	.07
$M1$	.48*	.51*	.52*	.53*	.47*
$M2$	.17	.19	.22	.25	.06
$M2.5$	.16	.19	.21	.24	.05
$M3$	.21	.26	.28	.30	.07
$M4$	.30	.33	.35	.37	.09
$M5$	.43[a]	.46[a]	.25[a]	.08	.06
1960–68					
Currency	.11	.11	.13	.15	.16
$M1$	.40	.42	.31	.18*	.07
$M2$	.42	.43	.33	.16*	.01
$M2.5$	.43	.44	.33	.17*	.02
$M3$	.40	.40	.27	.12	−.02
$M4$	.42	.43	.30	.15	−.02
$M5$	.46	.43	.44*	.34*	.14

* The asterisk signifies satisfaction of the dual criteria.

[a] These entries exhibit a negative sum of coefficients contrary to expectation.

the appropriate scale variable in the relationship.

If money is held mainly to facilitate transactions, the appropriate scale variable to use in the empirical definition of money should include financial and interindustry transactions as well as those associated with the purchase of final goods and services. If money demand is related to wealth in a portfolio balance model, wealth or permanent income would reflect the underlying assumptions more accurately than would current *GNP*. Thus, the wealth constraint would appear to be relevant to the more broadly defined monetary aggregates while the transactions motive applies more closely to the narrowly defined medium of exchange. In either case, *GNP* is not the best variable to use although it probably reflects transactions more closely than it reflects wealth. By this reasoning, the tests presented here and elsewhere in the literature may be biased against the more inclusive aggregates in the attempt to discover the appropriate definition of money.

The usefulness of the empirical definition of money from the policy point of view depends on the structure of the underlying model. In particular, the endogeneity of the money supply becomes particularly troublesome to policy makers if the best definition includes the liabilities of non-bank financial intermediaries. Given a model in which asset stocks are endogenous, policy multipliers can be obtained from either reduced form equations in which all exogenous variables are represented, or simulations of a structural model containing deposit supply as well as demand relationships. Discussions of the empirical definition of money represent neither alternative adequately, and hence their usefulness to the policy making process is limited. (See Ando and Modigliani, De Prano and Mayer, Friedman and Meiselman, and, more recently, de Leeuw and Kalchbrenner.)

One endogenous financial variable which has not been considered in the present exercise is the liquidity represented in the value of policy reserves of life insurance companies. Policy loans made at 5 percent have become important as market rates have climbed above that value. A regression of reserves less policy loans as a ratio to net worth on various variables clearly establishes this asset as a substitute for savings deposits and marketable assets.[3]

$$(R - PL)/NW = -\underset{(-2.1)}{.14}\,RCB - \underset{(-3.9)}{.43}\,ARSD - \underset{(-3.4)}{.11}\,ARCP + Z$$

$$R^2 = .997 \quad DW = .96 \quad DF = 45$$

The equation indicates the large interest rate effect from savings deposits, *ARSD*, as well as substitution with corporate bonds, *RCB*, and commercial paper, *ARCP*. The inclusion of endogenously supplied substitutes for financial assets affects the solution of the system and the magnitude of the reduced form coefficients. This "asset," which has been ignored until recently (see Hulett), should be considered in specifying a full financial model.

In summary, the Friedman-Meiselman-Kaufman exercise of empirical money definition may be questioned. Its results are not particularly stable, nor do they "explain" a large fraction of the variance in *GNP*. *GNP* is not the best scale variable, and its use may bias the results in favor of narrow definitions of money. Also, its usefulness for the exercise of policy founders on the two problems of endogeneity of certain included variables and omission of various exogenous variables needed for the reduced form model (see Laidler). If one wants to study asset substitutability, perhaps the best way is the frontal assault which has been attempted by various authors including Chetty, Gramlich and Hulett, and Hamburger.

REFERENCES

W. H. L. Anderson, *Corporate Finance and Fixed Investment,* Boston 1964.

A. Ando and F. Modigliani, "Velocity and the Investment Multiplier," *Amer. Econ. Rev.*, Sept. 1965, *55,* 693–728.

V. K. Chetty, "On Measuring The Nearness of Near-Moneys," *Amer. Econ. Rev.,* June 1969, *59,* 270–81.

F. de Leeuw and J. Kalchbrenner, "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilization—Comment," *Review,* Federal Reserve Bank of St. Louis, Apr. 1969, 6–11.

M. DePrano and T. Mayer, "Autonomous Expenditures and Money," *Amer. Econ. Rev.,* Sept. 1965, *55,* 729–52.

M. Friedman and D. Meiselman, "The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States, 1897–1958," in E. C. Brown et al., eds., *Stabilization Policies: Commission on Money and Credit,* Englewood Cliffs 1963.

M. Friedman and A. J. Schwartz, "Monetary Statistics of the United States," unpublished manuscript, Nat. Bur. Econ. Res.

H. Galper, "Alternative Interest Rates and The Demand For Money: Comment," *Amer. Econ. Rev.*, June 1969, *59,* 401–07.

E. M. Gramlich and D. T. Hulett, "Savings Flows, Mortgage Markets, and Residential Construction in the FRB-MIT Econometric Model," *Conference on Savings and Residential Financing,* U.S. Savings and Loan League, Oct. 1970.

M. Hamburger, "The Demand for Money by Households, Money Substitutes, and Mone-

---

[3] In this equation (from Gramlich and Hulett) $R-PL$ is life insurance company reserves less policy loans; NW is net worth of the public; RCB is the corporate bond rate; $ARSD$ is a three period Almon distributed lag on a weighted average of the rates on savings deposits and consumer *CD*s at commercial banks, deposits at mutual savings banks, and savings and loan shares; $ARCP$ is a distributed lag on the commercial paper rate; and Z represents terms in disposable income, capital gains, and the constant term.

tary Policy," *J. Polit. Econ.*, Dec. 1966, *74*, 600–23.

———, "Alternative Interest Rates and The Demand for Money: Comment," *Amer. Econ. Rev.*, June 1969, *59*, 407–12.

J. R. Hicks, *Value and Capital*, 2nd ed., London 1957.

D. Hulett, "The Mortgage Market: Alternative Specifications," mimeo. 1969.

J. Johnston, *Econometric Methods*, New York 1963.

G. Kaufman, "More on An Empirical Definition of Money," *Amer. Econ. Rev.*, Mar. 1969, *59*, 78–87.

D. Laidler, "The Definition of Money: Theoretical and Empirical Underpinnings," *J. Money, Credit, Banking*, Aug. 1969, *1*, 508–25.

T. H. Lee, "Alternative Interest Rates and the Demand for Money: The Empirical Evidence," *Amer. Econ. Rev.*, Dec. 1967, *57*, 1168–81.

Dependency Rates and Savings Rates: Comment

By KANHAYA L. GUPTA*

In a recent issue of this *Review*, Nathaniel Leff examined the role of demographic factors in the determination of aggregate savings rates, using international cross-section data. His major conclusion was "... that dependency ratios are a statistically distinct and quantitatively important influence on aggregate savings ratios, both for the 74 countries considered as a whole and within the subsets of developed and underdeveloped countries" (pp. 893–94). In this note, we shall present evidence showing that, contrary to Leff's findings, dependency ratios play an insignificant role in determining savings rates in the majority of the underdeveloped countries.

In a recent paper, I argued and showed that the treatment of underdeveloped countries as a single group is not very meaningful. In fact, very often such a treatment conceals more information than it reveals. It was then argued that a more satisfactory way is to subdivide these countries according to per capita income levels. Following the classification adopted in that paper, I divided the underdeveloped countries into three groups: (I) those with per capita income between \$0–124; (II) those between \$125–249; and (III) those between \$250–675.[1] Table 1 gives the identification of the countries in each group. Using Leff's data, we estimated his equations for each group separately and for the three groups combined together. The equations are:

TABLE 1—COUNTRIES IN EACH GROUP

Group I (9 Countries)	
India	Taiwan
Kenya	Tanganyika
Pakistan	Thailand
South Korea	Uganda
Sudan	
Group II (16 Countries)	
Ceylon	Jordan
Dominican Republic	Mauritius
Ecuador	Morocco
Egypt	Paraguay
Ghana	Peru
Honduras	Philippines
Iran	Portugal
Iraq	Tunisia
Group III (22 Countries)	
Argentina	Japan
Barbados	Malaysia
Brazil	Malta
British Guiana	Mexico
Bulgaria	Nicaragua
Chile	Panama
Costa Rica	Poland
Cyprus	Spain
El Salvador	Trinidad-Tobago
Greece	Turkey
Jamaica	Uruguay

$$(1) \quad S/Y = f_1(Y/N, g, D_1, D_2)$$

$$(2) \quad S/Y = f_2(Y/N, g, D_3)$$

$$(3) \quad S/N = f_3(Y/N, g, D_1, D_2)$$

$$(4) \quad S/N = f_4(Y/N, g, D_3)$$

Following Leff, all the equations were estimated in log-linear form. The symbols used stand for:

S/Y = aggregate savings ratio
S/N = per capita aggregate savings
Y/N = per capita income
g = rate of growth of per capita income
D_1 = percentage of population aged fourteen or less
D_2 = percentage of population aged sixty-five or more
D_3 = total dependency ratio, the sum of D_1 and D_2

In Table 2, we present the results of equations (1) and (3) only, because they are really the equations capturing the effects of

---

* Associate professor, University of Alberta. I am grateful to Leff for providing me with the data.

[1] The results of an alternative classification adopted in the United Nations Report were similar.

TABLE 2—RESULTS FOR DIFFERENT GROUPS

Dependent Variable	Intercept	$Ln\ Y/N$	$Ln\ g$	$Ln\ D_1$	$Ln\ D_2$	$\bar{R}^2$	F
Group I							
(1) $Ln\ S/Y$	3.9549	0.4548 (0.3768)	0.1263 (0.3737)	−0.7685 (0.1481)	−0.6475 (0.9446)	−0.29	0.55
(2) $Ln\ S/N$	−0.3564	1.6112 (1.6008)*	0.1309 (0.4644)	−1.0487 (0.2424)	−0.5585 (0.9770)	0.403	2.35
Group II							
(3) $Ln\ S/Y$	5.7068	−0.1372 (0.4052)	0.0234 (2.437)**	−0.6172 (0.7274)	−0.0449 (0.1419)	0.240	2.19
(4) $Ln\ S/N$	2.7274	0.8330 (2.6194)**	0.0243 (2.6937)**	−0.9724 (1.2200)	−0.1404 (0.4726)	0.495	4.67
Group III							
(5) $Ln\ S/Y$	14.2616	− 0.0584 (0.2401)	0.0347 (2.2646)**	−2.6974 (4.4973)***	−0.8866 (2.5979)***	0.564	7.78
(6) $Ln\ S/N$	8.9245	0.9172 (3.7661)***	0.0345 (2.2460)**	−2.4937 (4.1514)***	−0.7803 (2.2831)**	0.766	18.15
All Groups							
(7) $Ln\ S/Y$	9.3209	0.1624 (1.9444)**	0.0258 (3.1198)***	−1.8402 (4.5442)***	−0.5416 (2.700)***	0.471	11.22
(8) $Ln\ S/N$	4.6341	1.1501 (14.7254)***	0.0271 (3.5020)***	−1.8012 (4.7578)***	−0.5014 (2.7803)***	0.900	104.57

Notes: t values are given in the parentheses; * significant at 10% level; ** significant at 5% level; and *** significant at 1% level.

the demographic factors studied by Leff.[2] Looking at this table, we can draw the following inferences:

a) For Group I, the only significant variable is the level of per capita income. Furthermore, in terms of equation (2), even the combined effect of $ln\ D_1$ and $ln\ D_2$ fall short of the coefficient of $ln\ Y/N$. This is just the opposite of the results obtained by Leff for the underdeveloped countries.

b) For Group II, the results are slightly different in that now, in addition to the level of per capita income, the rate of growth of income is also significant. However, in so far as the effect of the dependency ratios is concerned, it still continues to be negligible.

c) For Group III, we get excellent results. Here all four variables are highly significant and their coefficients have the right signs. The effect of the dependency ratios is more important than that of per capita income. All these results are in full accord with Leff's findings. Looking at equations (7) and (8) in Table 2, we can observe that the results of Group III, are very similar to the results for the three groups combined together.

---

[2] The results of equations (2) and (4) are available from the author. The results of equation (3) test directly the hypothesis that the income elasticity of savings is significantly different from unity for the different groups. Since Leff also presents the results of this equation, it facilitates comparison of our results with his. We recognize that the $\bar{R}^2$'s for equation (3) will be biased, as pointed out by the referee and Leff, because of the presence of Y/N on both sides of equation (3). This follows from the fact that $S/N=(S/Y)(Y/N)$.

d) The income elasticity of savings is not significantly different from unity for any of the groups.

It is thus clear that the overall results for the underdeveloped countries merely reflect the savings behavior of countries in Group III. Dependency ratios do not appear to play any role in the other two groups. Hence the failure of the aggregate savings ratio to rise in these countries will have to be explained by factors other than the demographic factors considered by Leff. It should be noted here that Groups I and II account for more than 50 percent of the countries included in the underdeveloped category.

The above findings do not mean that Leff's hypothesis is completely wrong. It is possible that his hypothesis represents an *inter-group* relationship and not necessarily an *intra-group* relationshipfor all the groups.[3] While this interpretation leaves something of Leff's hypothesis intact, still it sheds very little light on the failure of the savings ratio to rise in the countries of Groups I and II.

A possible explanation of the failure of dependency ratios in the equations for Groups I and II is as follows. When income levels are as low as in these two groups, there is no margin left for savings. In fact, even in the absence of dependents these income levels would not provide the minimum level of living. Hence, it simply means that people are only sharing poverty—having two meals a day can lead to no more savings than only one meal a day. We suggest that demographic factors, like the dependency ratios, become operative and significant only when the per capita income of the *working population* reaches a level where it can provide more than a minimum level of living, thus generating potential savings. These potential savings can then be prevented from being realized by the presence of dependents.

REFERENCES

K. L. Gupta, "Foreign Capital and Domestic Savings: A Test of Haavelmo's Hypothesis with Cross-Country Data: A Comment," *Rev. Econ. Statist.*, May 1970, *52*, 214–15.

N. H. Leff, "Dependency Rates and Savings Rates," *Amer. Econ. Rev.*, Dec. 1969, *59*, 886–96.

United Nations, *Report on the World Social Situation*, New York 1961.

---

[3] The managing editor suggested this possibility and I am grateful to him for his suggestion.

Dependency Rates and Savings Rates: Comment

By NASSAU A. ADAMS*

In a recent paper in this *Review*, Nathaniel Leff concludes, on the basis of an analysis of cross-section data covering some seventy-four developed and underdeveloped countries, that demographic conditions are " . . . a major determinant of aggregate savings rates," that dependency ratios are " . . . a statistically distinct and quantitatively important influence on aggregate savings ratios, . . . " and that "High dependency ratios—and ultimately high birth rates—are among the important factors which account for the great disparity in aggregate savings rates between developed and underdeveloped countries" (pp. 893–94). In view of the central role which the aggregate savings rate has traditionally played in development theory, and in view furthermore of the growing concern in the international community with the question of population growth, the above conclusions linking these two issues clearly warrant attention.

The present note will examine the basis for Leff's conclusions at two levels. First, in terms of his theoretical rationale for expecting such a relationship; second, in terms of an analysis of his statistical results.

I

So far as theoretical rationale is concerned, the point is made that population growth typically means high birth rates and hence a high proportion of children in the population. And since "[children] . . . contribute to consumption but not to production, . . . [they] might be expected to impose a constraint on a society's potential for savings," (Leff, p. 887).

There seems to be two main weaknesses with this argument. First, it assumes that the factors that determine output are independent of the number of dependents for which provision has to be made. This need not be the case. It may well be that the pressure of increased family size on individual motivation and response favorably affects productivity and output. As Richard Easterlin puts it in a sobering discussion of the population issue, "Population pressure arising from mortality-reduction may provide the spur to work harder, search out information, increase capital formation, and try new methods" p. 104. This being so, the effect of higher population growth rates[1] may well be higher levels of output,[2] and perhaps higher levels of savings and capital formation as well to provide for the future needs of dependents, and at the abstract level of theorizing, it is not at all obvious that a negative relationship between these variables is to be expected.

The second weakness concerns the use of age fourteen as the cut-off point for dependent children. In the developed countries, or perhaps even in the cities of developing countries, this figure may well be appropriate, but it is totally inadequate in respect to farm families in the developing countries, most of whose children start to become economically useful at age six or seven. And since farm families account for the bulk of the population of the developing countries, the meaning of the analysis for these countries is clearly in doubt.[3]

---

* Economist, United Nations Conference on Trade and Development, New York. The views expressed are those of the author only. I am indebted to V. K. Sastry for helpful discussion of some of the issues raised in this note, as well as for the reference given in fn. 6.

[1] Resulting, e.g., from mortality reductions accompanied by unchanged high birth rates.

[2] And it may be noted here that among developing countries there is apparently a significant positive relationship between the rate of growth of output and the rate of population growth (see the article by John Conlisk and Donald Huddle).

[3] It would clearly be absurd to say that a farm child of ten years of age contributes to consumption but not to production. And since this proposition is the main basis for giving economic interpretation to the demographic variable (at least that involving child dependency), this is an important weakness in the analysis.

II

The equations Leff fitted to data for a single year, 1964, were of the form:

$$(1) \qquad S/Y = f(Y/N, g, D_1, D_2)$$

where

S/Y = the aggregate savings ratio
Y/N = per capita income
g = rate of growth of per capita income (annual average for the preceding five years)
D_1 = the proportion of population age fourteen and below
D_2 = the proportion of the population age sixty-five and above.

The rationale for introducing D_2 is similar to that for D_1, but it must be noted here that this variable has no bearing on the issue of population growth, high birth rate, etc., which is much emphasized by Leff in his conclusions.

These equations were fitted to data for seventy-four developed and developing countries grouped together and for forty-seven developing countries and for twenty western developed countries separately. Leff's results are summarized in Table 1. The dependency variables D_1 and D_2 appear significant in both the equation for all seventy-four countries as well as in that for the forty-seven developing countries, although in the latter the proportion of variation explained amounts to only 24 percent (compared with 57 percent for the all-country sample). These results provide the basis for Leff's conclusions.

There are several problems involved in Leff's interpretation of these results. The first point to note is that for the forty-seven developing countries, the dependency ratio D_1 varies only between 40–46 percent (and D_2 between 3–5 percent). It seems hardly convincing to argue that variations in the "dependency" ratio within a range of 40 and 46 percent could be a major factor accounting for differences in aggregate savings rate. But the important point here is that Leff presents no argument why the functional form

$$S/Y = f(Y/N, g, D_1, D_2),$$

is the theoretically correct one. In fact his arguments lead to the expectation that $S/Y = f(Y/N, g, D_1)$ should be just as valid, and indeed D_2 is mentioned almost as an afterthought in the development of the argument. The fact is, however, that using his own data for the forty-seven developing countries,[4] and fitting the equation $S/Y = f(Y/N, g, D_1)$, shows D_1 to be totally without statistical significance as an explanatory variable. The relevant results are summarized in Table 2. It appears from these results that when D_1 is introduced as the only dependency variable, it is without significance.

TABLE 1—REGRESSION COEFFICIENTS: (*ln* S/Y DEPENDENT VARIABLE)

	Independent Variables				Const	$\bar{R}^2$	F
	ln Y/N	*ln* g	*ln* D_1	*ln* D_2			
A	.1596 (2.8776)	.0254 (3.2792)	−1.3520 (4.6406)	− .3990 (2.5623)	7.3439 (5.7289)	.5697	25.1604
B	.1292 (1.8487)	.0227 (2.8079)	−1.2297 (2.7636)	− .4455 (2.1554)		.2419	4.6685
C	.0035 (.0296)	.2589 (1.6228)	− .4324 (1.7099)	− .4916 (2.6547)		.4395	4.7245

Note: *t*-ratios in brackets.
A: All 74 countries; B: 47 developing countries; C: 20 western developed countries.

---

[4] I am grateful to Leff for kindly making available to me his data on which these results are based.

TABLE 2—REGRESSION COEFFICIENTS, 47 DEVELOPING COUNTRIES

	Independent Variables					Const.	$\bar{R}^2$
$\ln S/Y$	$\ln Y/N$	$\ln g$	$\ln D_1$	$\ln D_2$	$\ln D_3$		
(1)	0.0687	0.0771	−0.45921			3.8889	0.1348
	(0.97)	(1.81)	(−1.40)			(2.74)	(0.13)
(2)	0.1251	0.0894		−0.668		1.9537	0.0993
	(1.64)	(1.97)		(−0.44)		(5.37)	(0.10)
(3)	0.0678	0.0798			−0.8036	5.2530	0.1485
	(0.98)	(1.90)			(−1.64)	(2.58)	(0.15)
(4)	0.1230	0.1134	−1.3753	−0.5462		7.7041	0.2289
	(1.75)	(2.66)	(−2.87)	(−2.50)		(3.79)	(0.23)

Similarly for D_2, as well as for D_3 (defined as D_1+D_2). However when D_1 and D_2 are introduced together as separate variables they become highly significant. But these results seem to be a mere statistical curiosity, since the data are evidently plagued by multicollinearity, D_1 and D_2 being highly correlated, as shown in the correlation matrix of Table 3.

TABLE 3—CORRELATION MATRIX, BASED ON LEFF'S DATA FOR 47 DEVELOPING COUNTRIES

(all data in logarithmic form)

	Y/N	g	D_1	D_2	D_3
S/Y	.28	.32	−.32	.17	−.33
Y/N		.19	−.42	.52	−.37
g			−.17	.35	−.12
D_1				−.81	.98
D_2					−.70

Multicollinearity is also indicated by the low proportion of variation explained by equation (4), i.e., 23 percent, compared with the high proportion of D_1 explained by D_2 and vice versa, i.e., 64 percent.[5] The instability of the coefficients, i.e., the fact that the coefficients for D_1 and D_2 change greatly when they are introduced together as against separately (compare equations (1) and (2) with equation (4)) only emphasizes further the problem of multicollinearity in the data. And since in the presence of multicollinearity the standard error of the regression coefficients might be quite misleading,[6] the fact that both D_1 and D_2 appear with low standard errors in equation (4) cannot be accepted as establishing the statistical significance of these variables.

It may therefore be concluded that the results shown in Leff's paper are statistically unreliable, and that it is therefore quite out of the question to conclude from them that dependency ratios, birth rates, population growth, etc., are important factors affecting the aggregate savings rate. More careful analysis of the data reveal that they permit no such conclusions to be drawn.[7]

III

The major conclusion reached by Leff, that "High dependency ratios—and ultimately high birth rates—are among the important factors which account for the great disparity in aggregate savings rates between developed and underdeveloped countries" can now be seen to be entirely unwarranted. The fact is that developed countries tend to have high savings rates and low birth rates, and that for the underdeveloped countries the situation tends to be the reverse. This

---

[5] Thus regression of D_1 against D_2 yields (in log form)

$$\ln D_1 = \underset{(82.45)}{4.1607} - \underset{(-9.11)}{0.3322} \ln D_2 \qquad \bar{R}^2 = 0.64$$

[6] For a discussion of the problem of interpreting standard errors in the context of multicollinearity, see the article by Richard Stone.

[7] It is in any event difficult to see how such strong conclusions could have been drawn from an equation which was only able to explain 23 percent of intercountry variations in the savings rate.

being the case, by grouping developing and developed countries together, birth rates (or the corresponding dependency ratios), will "explain" differences in savings rates between these groups of countries. In fact a dummy variable taking the value 1 for developed countries and 0 for developing countries will serve equally well, as the following results show.[8] (Z is the dummy variable):

$$\ln S/Y = \underset{(7.717)}{2.2826} + \underset{(.8961)}{0.0481} \ln Y/X + \underset{(2.6455)}{0.0986} \ln g + \underset{(4.3962)}{0.5019} Z \qquad (2)$$

$$\bar{R}^2 = 0.56$$

The results here closely parallel Leff's equation (3); the proportion of variation explained is approximately the same (56 percent), and the role played by the dummy variable corresponds to that of the dependency variable in his equation, the level of significance being approximately the same for both variables.[9]

From these results it is hardly possible to argue that the birth rate is an important factor affecting differences in savings rates. Clearly a host of factors determine birth rates, and these are evidently highly correlated with the level of development, per capita income, etc. A host of factors also determine savings rates, and these are also evidently correlated with the level of development, per capita income, etc. To group developed and developing countries together and show an inverse relationship between savings rates and birth rates hardly provides a basis on which to draw conclusion about economic phenomenon. And as we have seen, the relationship breaks down when the analysis is confined to the developing countries. And this does not seem to be merely a statistical phenomenon. As was argued earlier, any economic relationship between birth rates (or child dependency ratio) and the aggregate savings rate can only be a very tenuous one, and it would seem rather far-fetched to suppose that amidst the welter of factors that undoubtedly affect intercountry difference in aggregate savings, this could be strong enough to be statistically significant.

REFERENCES

J. Conlisk and D. Huddle, "Allocating Foreign Aid: An Appraisal of a Self-Help Model," *J. Develop. Stud.*, July 1969, *6*, 245–51.

R. A. Easterlin, "The Effects of Population Growth on the Economic Development of Developing Countries," in R. D. Lambert, ed., *The Annals of the American Academy of Political and Social Science*, Philadelphia 1967.

N. Leff, "Dependency Rates and Savings Rates," *Amer. Econ. Rev.*, Dec. 1969, *59*, 886–96.

R. Stone, "The Analysis of Market Demand," *J. Roy. Statist. Soc.*, 1945, *108*, pt. 3–4, 286–382.

---

[8] Based on Leff's data for seventy-four developed and underdeveloped countries.

[9] These results follow from the fact that developed and underdeveloped countries form two clusters, the dependency ratio (D_1) ranging between 24–31 percent for the former and for the latter between 40–46 percent.

Dependency Rates and Savings Rates: Reply

By NATHANIEL H. LEFF*

My paper presented data on two questions relating to the determinants of international savings rates. First, it showed that both in developed and in less developed countries, cross-section analysis did not indicate the level of per capita income to be an important determinant of S/Y. Second, it showed that both for the overall sample of seventy-four countries and within the subsets of more developed and less developed countries, dependency rates are a significant factor influencing aggregate savings rates. In their comments, Kanhaya Gupta and Nassau Adams raise a number of questions concerning these results, especially as they apply to the less developed countries. I will reply to their comments in sequence.

I

Gupta has grouped the observations for the less developed countries according to their level of per capita income, and estimated equations with $ln\ S/Y$ as the dependent variable.[1] His estimates indicate that for the two lowest income groups of countries, neither the level of per capita income nor the dependency rates is a significant determinant of aggregate savings rates. I doubt that this result is due to the explanation which Gupta proposes: that at low income levels, "no margin is left for savings." W. A. Lewis long ago pointed out some of the fallacies of this argument (p. 346 *ff*.). Indeed, one might expect that at low income levels, both the precautionary and the future-income motives for saving and investment would apply with special force.[2] Rather, I suspect that the results presented by Gupta for the countries of Groups I and II may reflect the estimation bias introduced in these small samples by errors in measurement, both in the dependent and independent variables. Such errors might be especially severe in the poorest countries because of the difficulties of estimating the income and savings generated in agriculture, which are proportionately larger in these countries.

In any case, the crux of Gupta's comment is that the overall results for the underdeveloped countries, which indicate the importance of the dependency variables, "merely reflects the savings behavior of countries in Group III." An obvious way to establish this proposition rigorously would be a test of covariance such as that proposed by Gregory Chow. Gupta presents no such results. However, I have computed the F-statistics to test for the equality of the regression coefficients as between the estimates for the forty-seven country set and the estimates for the countries of Groups I, II, and III. The F-ratios are, respectively: 0.72, 0.53, and 0.45. Since the critical value of F (5, 37) is 2.47 at the .05 level, we cannot reject the null hypothesis that the subsets are part of the same regression structure as the entire sample.[3] Thus, contrary to Gupta's assertion, these countries belong statistically to the same regression structure as the estimates for the entire set of less developed countries, and the overall estimates apply also to them.

---

* Associate professor in the Graduate School of Business, Columbia University. I have benefited from discussions with Jacob Merriweather, M. Ishaq Nadiri, and Maurice Wilkinson in the preparation of this reply. They bear no responsibility for its contents.

[1] I confine my discussion to the results which Gupta presents for equations with S/Y as the dependent variable because the equations with S/N as the dependent variable are derived from the S/Y equations, and contain no independent behavioral information. As noted in my paper, $S/N \equiv S/Y \cdot Y/N$. Consequently, given Y/N, S/N is determined by S/Y.

[2] It is also not at all certain that rates of time preference for countries vary inversely with levels of per capita income. See the discussion in Leff (pp. 618–20).

[3] The F-statistics cited in the text were computed from equations in which $ln\ Y/N$, $ln\ g$, $ln\ D_1$ and $ln\ D_2$ were specified as the independent variables. If $ln\ D_3$ is substituted for the other dependency variables, the results of the Chow test are similar. The F-statistics for the three groups are respectively: 1.10, 0.22, and 0.29. Similar results of statistically insignificant F ratios were obtained when the Chow test was applied to equations estimated with $ln\ S/N$ as the dependent variable.

II

Adams raises a number of other questions, both analytical and statistical. He states that it "seems hardly convincing to argue that variations in the dependency ratio within a range of 40 and 46 percent [in the less developed countries] could be a major factor accounting for differences in aggregate savings rates." That, however, is something to be settled after considering the data rather than asserted *ex cathedra*. He also suggests that I presented no theoretical justification for estimating equations in which both D_1, and D_2 are specified. The discussion on pages 887–89 of my paper, however, explains why both of these variables are theoretically relevant. Furthermore, as I will show below, his estimates of equations omitting one or the other of the dependency variables are biased by a serious error in specification.

Before discussing the statistical issues, however, let us consider Adams' analytical points. He suggests that the pressures of increased family size on individual motivation may raise productivity and output in less developed countries. This argument suffers from two weaknesses. First, it seems to imply that in the absence of larger family size, families will persist in sub-optimizing household allocational decisions. This goes against the abundant evidence which has accumulated concerning the economic rationality of micro-decision making in the less developed countries. One can, of course, argue the existence of household slack or leisure which, with larger family size, will be allocated to productive activities. We must distinguish, however, between the motivation and the capacity for such a response. In particular, as has often been noted, children are a time-intensive commodity (see Gary Becker). Apart from their simple consumption needs, the care of children involves large additional demands on household time and energies. Consequently, contrary to Adams' assertion, larger family size may have a debilitating effect, and *reduce* the capacity "to work harder, search out information, increase capital formation, and try new methods." At the least, higher dependency rates might be expected to lower the participation rate of women in the labor force.[4]

Adams also notes the difficulties involved in using age fourteen as the cut-off point for computing the dependency ratio D_1. To avoid some degree of arbitrariness in establishing a cut-off point, however, would require considerable knowledge concerning labor-force participation and worker productivity by age-cohort in the sample of seventy-four highly varied economies. In the absence of such data, I preferred to use a uniform procedure for all countries, and worked with the conventional demarcation point for the child dependency rate, age fourteen. Moreover, even if a fraction of the population aged fourteen or less does participate in the labor force in the less developed countries, their contribution to output is probably much less than the productivity of adults because of their inferior experience and strength.[5] It is striking that despite the obvious simplifications of the methodology and the rough nature of the data used, the effects of dependency rates on savings rates showed up so clearly in the equations estimated, both for the sample of seventy-four countries considered as a whole, and in the subsets of more developed and less developed countries.

III

Noting the high correlation between $ln\ D_1$ and $ln\ D_2$ for the set of forty-seven less developed countries, however, Adams suggests that because of multicollinearity, the results presented in my paper are a "mere statistical curiosity." We will examine below the possible bias introduced by multicol-

[4] The empirical evidence which Adams mentions from the study by John Conlisk and Donald Huddle, actually goes counter to his hypothesis. As discussed below, the correlation Adams cites indicates that in this sample, population growth was associated with a less than proportionate increase in the rate of growth of *per capita* output.

[5] Detailed data available for one less developed country, Mexico, do not indicate a high participation rate for the population aged fourteen or less. In 1960, the labor force participation rate for the cohort aged ten to fourteen years was only 15 percent. This relatively low rate obtained even though approximately 52 percent of the total labor force was in agriculture. See the study by El Colegio de Mexico (pp. 162, 175).

linearity. First, however, we should note that there is a straightforward way of clarifying this issue. In the equation where the total dependency burden, D_3 $(=D_1+D_2)$, is specified, the possible problem of collinearity which Adams raises does not arise. Consequently the estimates for equations specifying $ln\ D_3$ become of central importance in resolving the question of the effects of dependency conditions on savings rates.

My paper (p. 890) presented the results of such an estimate using the observations for all seventy-four countries. The regression coefficient for $ln\ D_3$ was -1.5, with a t-ratio of 4.9 and a *beta* weight twice as high as those for the other variables specified. Adams estimated a similar equation for the sample of forty-seven less developed countries. However, the results which he reports for this equation (equation (3) of his second table) present some difficulties in interpretation. Perhaps because of an imprecise computer program (see J. W. Langley), his results differ considerably from the estimates of the same equation using the same data, which Gupta reports.[6] They also differ from the estimates which I obtained using three different computer programs.[7]

Adams reports a regression coefficient for $ln\ D_3$ of $-.8036$, with a t-ratio of 1.64 and an $\bar{R}^2$ for the equation of .1485.[8] By contrast, estimating the same equation with the same data, Gupta reports a regression coefficient almost twice as large, -1.55, with a t-ratio of 3.71, and an $\bar{R}^2$ for the equation of .410. My results for the significance of the coefficient on $ln\ D_3$, obtained with three different computer programs, also indicated statistical significance above the .05 level. Thus the estimates for the equation in which $ln\ D_3$ was specified, and in which possible bias because of collinearity between $ln\ D_1$ and $ln\ D_2$ is not present, confirm the significance of the dependency burden as a factor accounting for the variance in the aggregate savings ratio in less developed countries.[9]

More generally, returning to the equations in which $ln\ D_1$ and $ln\ D_2$ were specified together, Adams seems to consider that the presence of a high simple correlation between two independent variables *ipso facto* results in valueless estimates because of multicollinearity. If multicollinearity biased the estimates of the equation in which $ln\ D_1$ and $ln\ D_2$ are specified together, however, we would expect that the estimates for these variables would lose their minimum variance properties, and the standard errors of these coefficients would increase. As seen in equation (4) of Adams' second table, however, precisely the opposite occurs: the t-ratios of the variables *rise* in the equation in which both $ln\ D_1$ and $ln\ D_2$ are specified together.

Finally, we should note that omission of the theoretically relevant, collinear variable *will* bias parameter estimates (Carl Christ, pp. 388–89, D. E. Farrar and R. F. Glauber, p. 106). This seems to be exactly what happened with Adams' equations in which he specified $ln\ D_1$ without $ln\ D_2$, and vice versa (equations (1) and (2) of his second table). This specification error (as well as the possible effects of imprecision in his computer program) may explain the relatively poor performance of Adams' equations in which $ln\ D_1$ and $ln\ D_2$ were employed separately.[10] Given the strong covariance between $ln\ D_1$ and $ln\ D_2$, it is not surprising that their coefficients change when one of these variables is omitted as compared with the equation when they are specified together.

---

[6] Gupta kindly made available his results for this equation.

[7] The computer programs which I used in estimating these equations were the NBER Stepwise Regression Program; MIAWOLS, written by M. Norman; and RAPE, for which the programming was done by W. J. Raduchel.

[8] A word on the critical t-values is in order here. Since we have definite theoretical expectations concerning the sign of the coefficient on the D variables, a one-tailed t test is appropriate. With 40 degrees of freedom, the critical values of t at the .05 and .1 significance levels are, respectively, 1.684 and 1.303.

[9] Adams may have been misled by an unreliable computer program. Still, he seems a bit cavalier in dismissing as "without significance" a coefficient which his own results indicate to have a t-ratio of 1.64, significant at a confidence level very close to .05.

[10] Adams indulges in something of an overstatement, however, when he says that when D_1 is specified without D_2, it is "totally without statistical significance as an explanatory variable." The t-ratio he presents (equation (1) of his second table) shows D_1 to be significant at the 10 percent level. As Kenneth Arrow has emphasized, dismissing results which are significant at this level may

IV

Adams also raises the issue of possible spurious correlation in the estimates relating to the sample for all seventy-four countries. That is, high dependency rates might be only a proxy variable for the general complex of conditions which comprise economic underdevelopment. I discussed this possibility in my paper (p. 891). To test for the possibility that the estimates showing the significance of the dependency variables were not simply picking up a "cluster effect," between the more developed and less developed countries, I estimated separate equations for the subsets of these countries. The results, presented in my paper, indicated the significance of the dependency variables within both subsets of countries.

In any case, the equation which Adams estimated with a dummy variable is hardly an adequate test for spurious correlation. A more appropriate way to test for the effects of dependency rates on savings rates—quite apart from the effects of the general complex of underdevelopment—would have been to specify the dependency variables *together* with the dummy variable to see if they retain their statistical significance. The results of an equation with this specification are presented below. A dummy variable was specified which took the value of 1 for the underdeveloped countries and 0 for the more developed countries. Student t-ratios are in parentheses, and the *beta* weights are underneath, in brackets.

$$(1)\quad \ln S/Y = \underset{(5.47)}{6.656} - \underset{(3.21)}{.398}\text{ dummy}$$

$$+\ \underset{\substack{(1.54)\\ [.205]}}{.083 \ln Y/N}$$

$$+\ \underset{\substack{(3.11)\\ [.233]}}{.023 \ln g} - \underset{\substack{(3.26)\\ [.573]}}{.971 \ln D_1} - \underset{\substack{(2.70)\\ [.505]}}{.394 \ln D_2}$$

$$\bar{R}^2 = .621 \qquad F \doteq 24.9$$

The t-value of the dummy variable is significant above the .01 level. This indicates that, not surprisingly, other features of underdevelopment aside from dependency conditions influence aggregate savings ratios. But the coefficients on $\ln D_1$ and $\ln D_2$ also retain their statistical significance above the .01 level. And while introduction of the dummy variable reduces the magnitude of the regression coefficient for $\ln D_1$ (and for $\ln Y/N$) as compared with the estimates (presented in my paper), for the equation without the dummy variable, the elasticities and the *beta* weights of the dependency variables continue to be very high.[11] Thus, contrary to Adams' assertion, equation (1) indicates that even when we control for other features of underdevelopment, dependency ratios are still important in accounting for the lower savings rates of the less developed countries.

V

Adams quotes a study showing a positive correlation between the rate of growth of output and the rate of population growth in developing countries. Adams might have noted, however, that the correlation he cites is between the rate of population increase and the rate of growth of *aggregate* rather than of per capita output (see Conlisk and Huddle, pp. 246–47). In fact, the elasticity of the aggregate growth rate with respect to

---

lead to Type I error and a costly loss in valuable information.

[11] Equation (1) was estimated with the observations for seventy-four countries, in order to permit comparison with Adams' results, which included all seventy-four countries. It is not clear, however, how to classify the observations for the seven Communist countries in this developed/underdeveloped dichotomy. Excluding the observations for those countries and reestimating equation (1) for the forty-seven less developed countries and the twenty developed countries gave the following results. The t-ratios are in parentheses and the *beta* weights are underneath, in brackets.

$$\ln S/Y = \underset{(4.95)}{6.417} - \underset{(1.98)}{.270}\text{ dummy} + \underset{\substack{(2.02)\\ [.411]}}{.118 \ln Y/N}$$

$$(1a)\qquad +\ \underset{\substack{(3.12)\\ [.183]}}{.023 \ln g} - \underset{\substack{(3.20)\\ [.444]}}{.988 \ln D_1} - \underset{\substack{(2.46)\\ [.367]}}{.378 \ln D_2}$$

$$\bar{R}^2 = .579 \qquad F = 19.2$$

This equation, which avoids the possible difficulties which may arise in the classification of the Communist countries, shows even more strikingly the effects of the dependency variable, even when the dummy variable is specified.

the rate of population increase in this sample was less than unity (0.69).[12] This result indicates that for this sample, higher population growth was associated with a *less* than proportional rate of increase in the growth of per capita *GNP*.

In any case, my paper related to savings rates rather than growth rates. The results implied that if it were not for the high dependency rates induced by high birth rates, less developed countries might be able to achieve higher savings ratios.[13] This might make possible higher rates of investment and growth than they would otherwise attain. In addition, they could finance a larger share of their investment with domestic savings rather than be dependent on foreign capital inflow, with its sometimes unwelcome side effects (see K. B. Griffin and J. L. Enos). Some views of development indeed stress precisely the capacity for autonomous economic progress, independent of the need for foreign resource flows. In this context, the term "dependency rates" takes on new, and broader meaning—dependency vis-à-vis the capital-exporting countries.

REFERENCES

K. J. Arrow, "Decision Theory and the Choice of a Level of Significance for the *t*-test," in I. Olkin et al., eds., *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling,* Stanford 1960.

G. S. Becker, "A Theory of the Allocation of Time," *Econ. J.*, Sept. 1965, *75,* 493–517.

H. B. Chenery, "Targets for Development," Columbia University Conference on International Economic Development, New York 1970.

G. C. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica,* July 1960, *28,* 591–605.

C. F. Christ, *Econometric Models and Methods,* New York 1966.

J. Conlisk and D. Huddle, "Allocating Foreign Aid: An Appraisal of a Self-Help Model" *J. Develop. Stud.* July 1969, *6,* 245–51.

L. E. Davis and R. E. Gallman, "The Share of Savings and Investment in Gross National Product during the 19th Century, United States of America," paper presented to the Fourth Congress of The International Economic History Association 1969.

D. E. Farrar and R. F. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited," *Rev. Econ. Statist.,* Feb. 1967, *49,* 92–107.

K. B. Griffin and J. L. Enos, "Foreign Assistance: Objectives and Consequences," *Econ. Develop. Cult. Change,* Apr. 1970, *18,* 313–27.

A. C. Kelley, "Demographic Cycles and Economic Growth: The Long Swing Reconsidered," *J. Econ. Hist.,* Dec. 1969, *29,* 633–56.

K. A. Kennedy and B. R. Dowling, "The Determinants of Personal Savings in Ireland: An Econometric Inquiry," *Econ. Soc. Rev.,* Oct. 1970, *1,* 19–51.

J. W. Langley, "An Appraisal of Least Squares Programs for the Electronic Computer from the Point of View of the User," *J. Amer. Statist. Ass.,* July 1967, *61,* 324–39.

N. H. Leff, "Marginal Savings Rates in the Development Process: The Brazilian Experience," *Econ. J.,* Sept. 1968, *78,* 610–23.

W. A. Lewis, *The Theory of Economic Growth,* Homewood 1955.

S. Robinson, "Aggregate Production Functions and Growth Models in Economic Development: A Cross-Section Study," unpublished doctoral dissertation, Harvard Univ. 1969.

T. P. Schultz, *A Family Planning Hypothesis and Some Empirical Evidence from Puerto Rico,* The RAND Corporation, Santa Monica 1967.

El Colegio de México, *Dinamica de la Población de México,* México 1970.

---

[12] The elasticities estimated in another study are somewhat lower, ranging from .35 to .55. (See S. Robinson as cited in Appendix Table 3 of Hollis Chenery.)

[13] My study was confined to a cross-section analysis. However, Allen Kelley's time-series country study of Australia (pp. 640–44) has also demonstrated an inverse relation between child dependency rates and personal saving. See also the results of a time-series study of personal savings in Ireland, by R. A. Kennedy and B. R. Dowling. Finally, in their discussion of the rise of the savings rate in the nineteenth-century United States, L. E. Davis and R. E. Gallman (pp. 21–22) also emphasize the importance of shifts in the dependency ratio.

A Property of a Closed Linear Model of Production: Note

By DAVID LEVHARI*

In a recent communication in this *Review*, Thomas Finn asserts that in a sufficiently well-behaved von Neumann model of production, the efficiency frontier tends to "flatten-out" as time goes on. He presents a graphic proof of this theorem. This proof is incomplete. The gap in the proof is the argument on page 534 of Finn's communication, " . . . the sequence of points a_1, $a_2, \ldots,$ cannot include a point that has coordinates greater than those of a^* (i.e., a_t cannot be farther than a^* from the origin) because we know that $\lambda^{(1-t)}F_t$ cannot be convex to the origin. Therefore we know that the distance between a_t and a^* converges to zero as t approaches infinity." The fact that a^* is an upper bound of the monotonic sequence of points a_t does not prove that a^* is the limit of this sequence. A monotonically increasing bounded sequence has the least upper bound as its limit but it has not been proved that a^* is, in fact, the least upper bound. I have tried to find a counter example that would fulfill all of Finn's assumptions and particularly his assumption that distinct points on the efficiency frontier F_2 must come from distinct input bundles. I was unable to find one that fulfills this assumption and thus it is still unknown to me whether the theorem stated is right or wrong.

In the following comment I would like to discuss the validity of the theorem for the general Closed Linear Model of Production. The flattening out property of successive transformation or efficiency frontier curves has been proved by E. Drandakis for the case of no joint production. The assumption "no joint production" is sufficient to assure that the asymptotic frontier is a hyperplane in any closed convex linear model in which the von Neumann ray is unique and the turnpike theorem holds. This is also discussed by J. R. Hicks in "The Story of a Mare's Nest" and there he poses the problem "to find out how much would be left of it (the flattening out) if the 'no joint supply' assumption were partially relaxed." An answer to this question is not given in the present comment and one still wonders whether non-joint production is also a necessary condition for the flattening to occur.

In the present note we support Hicks' assertion that the theorem does not hold for the general case of joint production and we show that there exists one class of well-behaved closed von Neumann models of production—interesting in their own right—where this property does not hold (however, they do not possess the property that distinct points on the efficiency frontier come from distinct input bundles). In this example the normalized efficiency frontiers in successive time periods coincide with one another. One wonders to what extent the example given below is exceptional. To answer this question, one would have to give a complete characterization of those dynamic models which do exhibit the property that their efficiency frontiers (in successive time periods) tend to hyperplanes. Here, a characterization will not be attempted.

Denote by S_t the vector of outputs produced at time t using the inputs S_{t-1} at time $t-1$.

The transformation of inputs into outputs is assumed to obey the following simple relation: $g(S_t)=f(S_{t-1})$ where g and f are homogeneous of the first degree and have the appropriate convexity-concavity properties (i.e., f is concave and g is convex). Both f and g are assumed differentiable. This is a completely well-behaved case.

The conditions for intertemporal efficiency are derived by maximizing one component of S_{t+1} given S_{t-1} and the rest of the components of S_{t+1}, that is by maximizing $g(S_{t+1})=f(S_t)$ subject to $g(S_t)=f(S_{t-1})$. Using Lagrange multipliers, we get that the condition for this is $f_{S_t^i}+\mu g_{S_t^i}=0$, $i=1, \ldots,$

---

* The Hebrew University of Jerusalem.

n where S_t^i is the ith component of S_t. By division:

$$\frac{f_{s_t^i}}{f_{s_t^j}} = \frac{g_{s_t^i}}{g_{s_t^j}} \qquad i, j = i, \ldots, n$$

This is the familiar condition that the marginal rate of substitution of inputs should equal their marginal rate of transformation as outputs. Using the homogeneity of f and g these equations are just sufficient to determine the unique ray of efficiency. As one expects, this ray is the von Neumann ray. In this case, we have the special feature that the economy is always on the turnpike (except in the initial and final stages).[1] Since we always start on the same ray, the shape of the normalized efficiency frontiers is always $g(S_t) = C$ where C is some constant depending on f for all t. For example, if the relation is

$$\sqrt{(S_t^1)^2 + (S_t^2)^2} = (S_{t-1}^1)^{\alpha} + (S_{t-1}^2)^{1-\alpha}$$

the shape of the efficiency frontier will be a quarter of a circle for all t.

It is still a puzzle to me whether Finn's theorem can be shown to be wrong also in other cases where the efficiency frontiers do not overlap.

---

[1] To see that this is really the Von Neumann ray' assume a balanced growth situation;

$$S_t = \lambda S_{t-1} \quad \text{or} \quad \lambda = \frac{f(S_t)}{g(s_t)}$$

Maximizing λ, we obtain:

$$\frac{f_{s_t^i}}{g_{s_t^i}} = \frac{f}{g} \qquad i = 1, \ldots, n$$

Using again the homogenity of the functions involved, we find that we have $n-1$ independent equations determing the ratios of the S_t^i or the ray in which:

$$\frac{f_{s_t^i}}{g_{s_t^i}} = \frac{f_{s_t^j}}{g_{s_t^i}}$$

This obviously describes the same rays as the equation in the text. Thus the condition of intertemporal efficiency implies that the economy is always on the von Neumann ray.

REFERENCES

E. Drandakis, *On Some Efficiency Properties of the Two Sectors Production Model,* Athens 1966.

T. J. Finn, "A Graphical Proof of a Property of a Closed Linear Model of Production," *Amer. Econ. Rev.,* June 1967, *57,* 531–35.

J. R. Hicks, "Prices and the Turnpike. The Story of a Mare's Nest," *Rev. Econ. Stud.,* Feb. 1961, *28,* 77–88.

Economics of Production from Natural Resources: Comment

By RICHARD F. FULLENBAUM, ERNEST W. CARLSON, AND FREDERICK W. BELL*

In a recent issue of this *Review*, Vernon L. Smith advanced a general model of the economic aspects of production from natural resources. Unfortunately, Smith makes a number of unsound assumptions and methodological errors in formulating his general model of natural resource exploitation. We shall show, using commercial fishing as a particular example, that the structure of Smith's model leads to absurd conclusions. In addition, we shall develop the correct specification for the competitive recovery of a fishery resource.

I. Misspecification in the Smith Model

Smith attempts to develop his theory of production from natural resources by specifying a cost function. In the case of commercial fishing, Smith states that . . . "The most natural general hypothesis about total operating cost for the individual fisherman requires it to be an increasing function of the vessel's catch rate, x, but a decreasing function of fish population, i.e., $\phi=\phi(x, X)$, with $\phi_1\equiv\partial\phi/\partial x>0$, and $\phi_2\equiv\partial\phi/\partial X<0$" (p. 413). He then formulates the following cost function for the fishing firm,

$$C = \phi(x, X, K) + \hat{\pi} \tag{1}$$

where C = total cost per unit of time; x = catch rate per vessel per unit of time; X = fish population; K = both the number of *homogeneous* firms (vessels) and a measure of the real capital stock in the industry; $\hat{\pi}$ = minimum return necessary to hold the vessel in the industry. As we shall see, the specification of (1) leads to unreasonable results. Actually, *total* cost per unit of time for any fishing firm is not a direct function of x, X, and K. Rather, total cost per firm per unit of time in long-run equilibrium is merely dependent upon factor prices. That is, total cost for the fishing firm *in the long run* is not explicitly influenced by output or x.[1] This is true since the output or catch rate per vessel, x, is determined by stock and technological externalities, e.g., $x=g(X, K)$. Assuming $\hat{\pi}$ represents the rate of return necessary to hold all factors in the fishing firm, the total cost *per unit of time* for the firm is simply equal to $\hat{\pi}$. Total long-run *industry* cost is then equal to $K\hat{\pi}$. Thus, average cost per unit of output, AC, for both firm and industry is given by,

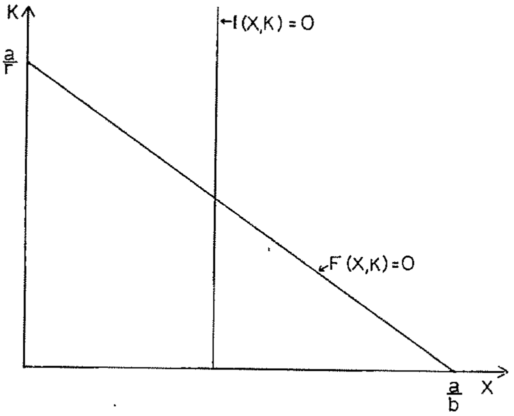


FIGURE 1. EXPLOITATION

$$AC \equiv \frac{K\hat{\pi}}{Kx} \equiv \frac{\hat{\pi}}{x} = \frac{\hat{\pi}}{g(X, K)} \tag{2}$$

Smith has unknowingly introduced a tech-

---

* The authors are economists with the Division of Economic Research, Bureau of Commercial Fisheries, U.S. Department of the Interior. The comments in this article do not necessarily represent the official position of the U.S. Department of the Interior.

[1] For example, the northern lobsterman (or any other fisherman) will make the same expenditures per unit of time (i.e., fuel, bait, labor, interest, and insurance costs) regardless of the number of lobsters found in his traps. The number of lobsters caught depends on aggregate fishing effort which is beyond the control of the lobsterman. *When we double the number of lobstermen exploiting a given fishery per unit of time, the total cost per unit of time is doubled, irrespective of what happens to catch rates.*

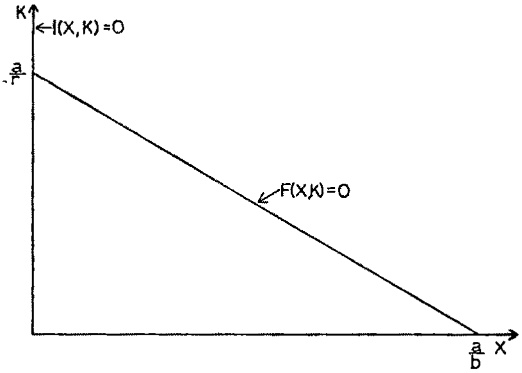


FIGURE 2. EXTINCTION

nological externality in the form of a market or pecuniary externality.[2] The basis of this misspecification stems from an apparent confusion between a short-run and long-run model of fishery exploitation. In the long run, any adjustments in the industry will be made in the number of *homogeneous operating units*, and not in the number of days fished per boat. However, a change in the number of units of time fished *does not change total cost per vessel* per unit of time fished. Secondly, if there is a variability in factor inputs per firm per unit of time—nowhere mentioned by Smith—then K is no longer a unique measure of the number of homogeneous vessels. His model must be corrected to eliminate these methodological errors since, as we shall see, they lead to untenable results.

II. Competitive Recovery of a Fishery Resource

In this section, we shall present the classical system of commercial fishing behavior developed by Anthony Scott, H. Scott Gordon, J. A. Crutchfield and Arnold Zellner, and illustrate its usefulness by a quadratic example. At various points, comparisons will be made to show the erroneous nature of Smith's model. The general model (3)–(8) for the commercial fishing industry may be expressed as follows:[3]

$$(3)\quad \frac{dX}{dt} = f(X, Kx)$$

$$(4)\quad Kx = Kg(X, K)$$

$$(5)\quad KC = K\hat{\pi}$$

$$(6)\quad \pi = pKx - KC = pKg(X, K) - K\hat{\pi}$$

$$(7)\quad \frac{dK}{dt} = \begin{cases} \delta_1\pi, & \text{if } \pi \geq 0 \\ \delta_2\pi, & \text{if } \pi < 0 \end{cases}$$

where p = price, π = pure profit, δ_1 and δ_2 are the rates of entry and exit of capital, respectively. All other variables are defined above. In the above system, (3) is the biological growth function, (4) is the industry production function, (5) is the industry cost function, (6) is the industry profit function, and (7) is the industry entry function. The equilibrium condition for the industry ($\pi = 0$) may be formulated as follows:

$$(8)\quad p = \frac{R(Kx)}{Kx} = \frac{\hat{\pi}}{g(X, K)}$$

$R(Kx)$ is the revenue function for the industry. As shown in (4), the only way the variables X and K enter the profit function is through their respective impact upon the vessel catch rate, x.

Let us now specify a quadratic example of the classical system which assumes no crowding externalities, i.e., $\partial x/\partial K \equiv g_2 \equiv 0$. First we have a quadratic growth function or

---

[2] Put differently, our objection to Smith's model revolves about his cost specification and the manner in which he introduces the technological externality. In his system, the marginal cost of K (i.e., marginal cost of additional factors) is given by

$$(1')\quad \frac{\partial KC}{\partial K} \equiv \frac{\partial}{\partial K}[K\phi(x, K, X) + K\hat{\pi}] = \frac{K\partial\phi}{\partial K} + \phi + \hat{\pi}$$

However, the properly specified marginal cost function for factors is

$$(2')\quad \frac{\partial KC}{\partial K} \equiv \frac{\partial}{\partial K}(K\hat{\pi}) = \hat{\pi}$$

In (1′), the marginal cost of K depends upon K. However, the relationship between total cost and K is strictly linear and homogeneous, i.e., the opportunity cost of inputs is assumed unaffected by the expansion of the exploiting industry, and thus, the marginal cost of inputs is constant.

[3] The more general model should include mesh considerations. Since Smith does not deal with this in his article, we shall abstract from mesh considerations in formulating our model.

(9) $$\frac{dX}{dt} = aX - bX^2$$

With exploitation of the fishery population by man, (9) becomes

(10) $$\frac{dX}{dt} = aX - bX^2 - Kx$$

Thus (10) is a particular specification of (3). Under a steady state assumption (i.e., equilibrium between population growth and man-made mortality), $dX/dt = 0$ and (10) becomes,

(11) $$Kx = aX - bX^2$$

Next, let us assume that the rate of loss to the population because of fishing is proportional to the number of firms (rK).[4] Multiplying the rate of loss because of fishing times the population, we have an industry production function or

(12) $$Kx = rKX$$

Substituting (12) into (11) and solving for K in terms of X, we have

(13) $$K = \frac{1}{r}(a - bX)$$

Therefore, the steady state relationship between K and X in (13) is inverse and linear. While linearity is not absolutely necessary, the following is, $dK/dX < 0$.

The negative relationship between K and X is reasonable since increases in fishing effort or K will, *celeris parabus*, always reduce the size of the stock. The inverse relation has been theoretically developed by M. Graham, M. B. Schaefer, T. I. Baranov, and R. J. H. Beverton and S. V. Holt and empirically verified by Schaefer, Beverton and Holt, and L. Van Meir. The inverse relation between X and K is one of the most fundamental laws of fishery population dynamics. Smith's Figure 3a is erroneous since it shows a positive relation for $F(X, K) = 0$ between K and X below some value of X. *This is logical and biological nonsense.* Increasing the number of vessels, or K, cannot increase the size of the biomass *in any range*. Smith's error is directly traceable to the misspecification of the cost function and its use in deriving the relation between X and K.[5]

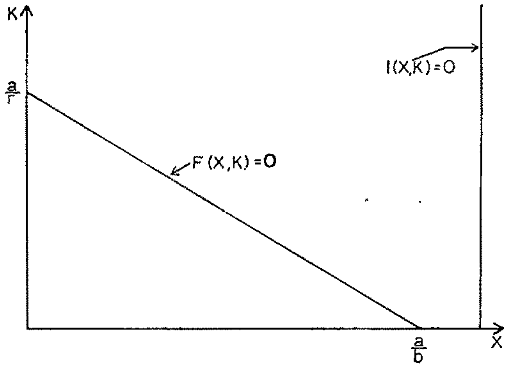


FIGURE 3. NON-EXPLOITATION

Assume a total revenue function (as does Smith) for the industry as follows (i.e., constant selling price),

(14) $$R(Kx) = pKx$$

and a properly specified total cost function for the industry,

(15) $$KC = K\hat{\pi}$$

or the number of vessels multiplied by their opportunity cost, $\hat{\pi}$. The profit function becomes

(16) $$\pi = pKx - K\hat{\pi}$$

---

[4] This implies a linearly homogeneous production function, holding the biomass constant. However, the use of an industry production function, which reflects crowding externalities and thus decreasing returns with respect to the number of boats, e.g.,

(3′) $$Kx = rK^{\alpha}X, \qquad 0 < \alpha < 1$$

would not change any of our conclusions and would even be more realistic. Nonetheless, (12) was used because of its traditional role in the fishery literature.

[5] In particular, one could derive the curve $F(X,K) = 0$ in the following manner. Assume a specified profit function for the fishing firm which has in it the type of cost function used by Smith. Taking the first partial with respect to x and setting equal to zero, one may derive an expression for x in terms of K and X, say $x = w(X,K)$. Then, under the assumption that $dX/dt = 0$, we have $0 = f(X) - Kw(X,K)$. The latter then is the implicit function $F(X,K) = 0$, which may be solved in terms of K as a function of X.

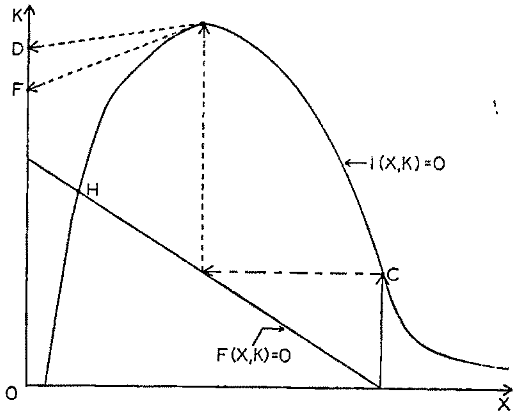


FIGURE 4

Competitive equilibrium is defined as a zero profit position or

$$0 = \bar{p}Kx - K\hat{\pi} \tag{17}$$

Since $Kx = rKX$, we may substitute in (17) and derive the equilibrium population with zero profit

$$X = \frac{\hat{\pi}}{\bar{p}r} \tag{18}$$

Equation (18) has the important property that the limit as $\bar{p}$ approaches ∞, $X \rightarrow 0$, and as price approaches 0, X approaches its natural upper limit. In (18) the population, X, increases with increases in the supply parameter $\hat{\pi}$ or fishing costs. When (18) is combined with (13), we can simulate 1) exploitation of the fishery resource, 2) nonexploitation, and 3) extinction as a lower limit to equation (18). Which of these three cases obtains depends upon the parameters $\bar{p}$, $\hat{\pi}$, and r. These cases are shown in Figures 1–3.

The classical model can be adjusted to show extinction as a dynamic result. Assume that price is not parametric to the individual fishery, but rather that p is a linear function of Kx (quantity), so that total revenue is quadratic in Kx. Figure 4 shows one possible set of curves and possible dynamic path. If we start at C and follow the dashed line, extinction (point D or F) is possible, even though point H is a static equilibrium position.

III. Comparative Policy Implications

Because of the nature of Smith's cost function, he reaches some absurd conclusions in his discussion of sole ownership versus competition. Let us assume that price is constant to the individual fishery. Assume further that there are no crowding externalities, i.e., $C_3 \equiv 0$. Thus, Smith postulates the Lagrangian function for the sole owner as

$$\psi = \bar{p}Kx - KC(x, X, K) + \lambda f(X, Kx) \tag{19}$$

Differentiating (19) with respect to K, X, and x, we find that

$$\bar{p} = C_1 + \lambda \tag{20}$$

where $\lambda = KC_2/f'$. If we want to obtain the sole ownership result, i.e., social marginal cost pricing, within a competitive free-access situation, a per unit of catch extraction fee, $U = \lambda$, should be imposed upon each fisherman. Under the assumptions of Smith's model $C_2 < 0$, and to the right of a population consistent with maximum sustainable yield, $f' < 0$. While Smith "proves" that under sole ownership, profit maximization guarantees exploitation of the biomass at the level at which $f' < 0$, the latter result does not necessarily follow in perfect competition. Suppose that with free exploitation, we are operating at X^*, the population at which total cost is equal to total revenue and that $\overline{X} < X^* < X^0$. It then follows that $f' > 0$, and $\lambda < 0$. At worst this implies a subsidy to a fishery which from an economic and biological point of view has been too intensively exploited.[6] At best, it demonstrates the lack of generality and lack of flexibility of Smith's "general" model.[7] On the other hand, the classical theory can define a tax which depends upon the parameters but not the slope of the biological growth function. For example, assume sole ownership and the profit function as defined in (16). Maximizing (16) subject to the constraint that $dX/dt=0$, is identical to taking the inverse of (13), $K^{-1}(X)$, substituting in (12), finding Kx in terms of K alone, and then substituting back in (16). Thus, the profit function is specified solely in terms of K:

---

[6] Overexploitation from a biological point of view usually refers to the reduction of the fishery population below that consistent with maximum sustainable yield.

[7] An anonymous referee has pointed out that 1) when $f' > 0$, Smith's tax does not apply because a sufficient condition for profit maximization for the sole owner requires $f' < 0$; and 2) that if $f' > 0$ in a purely competitive situation it is necessary to reduce fishing until the biomass is restored to its optimal level with $f' < 0$ and then to levy the tax mentioned by Smith in order to internalize social cost. Three comments seem appropriate. First, irrespective of the sign of f', a tax will reduce fishing effort! Thus, the imposition of a tax does not depend upon the second-order conditions for

$$\pi = \frac{\bar{p}arK}{b} - \frac{\bar{p}r^2K^2}{b} - K\hat{\pi} \tag{21}$$

Differentiating (21) with respect to K, setting equal to zero, we find

$$K^* = \frac{a}{2r} - \frac{\hat{\pi}b}{2\bar{p}r^2} \tag{22}$$

K^* reflects the number of boats per unit of time consistent with marginal cost pricing.[8] Within the context of a purely competitive, free-access situation a tax per boat (T_κ) may be imposed so as to attain K^*:

$$\begin{aligned} T_\kappa &= \frac{\bar{p}K^*x^* - K^*\hat{\pi}}{K^*} \equiv \bar{p}x^* - \hat{\pi} \\ &= \bar{p}\,\frac{ar}{2b} - \frac{\hat{\pi}}{2} \end{aligned} \tag{24}$$

When (24) is added to $\hat{\pi}$ as an element of cost per firm, equation (21) becomes:

$$\pi' = \frac{\bar{p}arK}{b} - \frac{\bar{p}r^2K^2}{b} - K(\hat{\pi} + T_\kappa) \tag{25}$$

Setting (25) equal to zero and solving for K we obtain the same solution as in (22). The tax per vessel does not depend upon the slope of the biological yield function; rather, it is positively related to the demand and supply parameters, $\bar{p}$ and r, respectively, and inversely dependent upon the opportunity cost per vessel. In addition, the tax, given the type of externalities assumed, will always be positive.

Thus, we have shown that the classical theory of commercial fishing is still on a firm foundation.

REFERENCES

T. I. Baranov, "On the Question of the Biological Basis of Fisheries," *Nauch. isseldov. Inst. Izv.*, 1918, *1*, No. 1, 81–128.

R. J. H. Beverton and S. V. Holt, "On the Dynamics of Exploited Fish Population," Ministry of Agriculture, Fisheries and Food, *Fishery Investigations*, 1957, Ser. II, *19*, 1–533.

J. A. Crutchfield and A. Zellner, Economic Aspects of the Pacific Halibut Fishery, *Fishery Industrial Research*, Apr. 1962, *1*, No. 1, Washington.

H. S. Gordon, "The Economic Theory of a Common-Property Resource: The Fishery," *J. Polit. Econ.*, Apr. 1954, *62*, 124–42.

M. Graham, "Modern Theory of Exploiting a Fishery and Application to North Sea Trawlers," *J. Conseil Int. Explor. Mer.*, July 1935, *10*, 264–74.

M. B. Schaefer, "Some Aspects of the Dynamics of Populations Important to the Management of the Commercial Marine Fisheries," *Inter-Amer. Tropical Tuna Comm., Bull.*, 1954, *1*, *2*, 25–56.

A. Scott, "The Fishery: The Objectives of Sole Ownership," *J. Polit. Econ.*, Apr. 1955, *63*, 116–24.

V. L. Smith, "The Economics of Production from Natural Resources," *Amer. Econ. Rev.*, June 1968, *58*, 409–31.

L. Van Meir, "An Economic Analysis of Policy Alternatives for Managing the Georges Bank Haddock Fishery," unpublished doctoral dissertation, Univ. Kan. 1969.

---

a profit maximum. Secondly, the restriction of fishing per se is a necessary and sufficient condition for a social optimum. If the optimal biomass, X^*, is attained through the restriction of fishing, then the optimal fishery catch K^*x^*, will also be reached and there is no need for a tax. Thirdly, within the structure of Smith's model *it is possible* that somewhere beyond *MSY* the amount of real capital exploiting a fishery is smaller under pure competition than under sole ownership. Then, the model requires an increase in the number of vessels in order to attain the sole ownership result; however, dynamically, this would move the biomass further away, not closer, to its optimal level! Thus, in any event, Smith's tax is meaningless.

[8] The solution in (22) may also be obtained by setting price (p) equal to marginal cost, so that:

$$\bar{p} = MC \equiv \frac{\hat{\Pi}}{\frac{\partial Kx}{\partial K}} \tag{4'}$$

This yields the identical expression for K^*.

Economics of Production from Natural Resources: Reply

By VERNON L. SMITH*

One always writes a paper on the assumption that a certain background knowledge of principles is shared by the reader. When this is not the case, communication can become very protracted, and when there is the kind of imperviousness to understanding that is revealed in the comment by Richard Fullenbaum, Ernest Carlson, and Frederick Bell (FCB), communication may be impossible. But I will do what I can in this reply to explicate those aspects of my original paper, with which FCB are having difficulties.

I

When I write a cost function $C=\phi(x, X, K)+\hat{\pi}$, I take it for granted that the reader is familiar with the textbook theory of least-cost production under certainty, and will know without instruction that output, x, is variable only by virtue of input variability (FCB note that I nowhere mentioned input variability). In the context of commercial fishing, it is therefore implicit in $\phi(x, X, K)$ that the firm has a production function $x=G(z; X, K)$ where z is a vector of inputs, and the technological externality parameters X (fish stock) and K (number of firms, if there is harvest "crowding") are postulated beyond the control of the individual firm when the resource is unappropriated. As a least-cost producer the firm chooses z so as to minimize $C=wz$ (where w is a vector of input prices), subject to the constraint G, given (X, K). Under the usual regularity conditions, this process defines a set of theoretical input demand functions and, therefore, a lowest possible cost function $\phi(x; X, K, w)$. In general, the resulting cost is variable with output, and the parameters (X, K, w) hypothesized to be outside the firm's control.

I chose to simplify the presentation of the model by assuming that the K firms were identical. This need not mean, as believed by FCB, that the firm cannot change its size, output, and structure of operations in choosing among different input combinations (this technological variability is what is specified by G), but merely that all firms are identical and make the same changes. An example of a "firm" (which is a unit of organization) might be a vessel, complete with nets or traps, or it might be a mother ship with satellite fishing boats. These, and other inputs, are in general assumed to be variable, if not in the short run, then in the long run. The size and composition of vessels, gear, labor applied, fuel consumed, frequency with which traps are run, or days and lengths of days spent fishing are all in general, assumed not to be invariable and unresponsive to input prices, the biomass of fish and the number of fishing units (where crowding or gear congestion is important), but the degenerate case of no variability is not excluded. In the case of crowding, to be sure the size and structure of fishing units may be important as well as the number so that K may not capture all that is essential in crowding.[1] But, as I took pains to note, neither may X capture all that is important about the stock of fish, whose size (or age)

---

[1] The number of firms, K, would also serve as a measure of the stock of capital only in the case of the example used in my original article, i.e., where a "firm" corresponds to a fixed "vessel." It is obvious that, except at a stationary equilibrium point, K would not also serve as an index of capital in cases where the development of the industry through time involved significant redesign of gear and vessels. In fact in an industry subject to wide changes in the stock of fish, one would expect flexibility in the design of vessels to command a premium so that the basic capital plant could accomodate wide variation in operating conditions (and even in the species harvested) by adjustments in operating inputs. My model would certainly require reinterpretation, and, I hope, would encourage reformulation, to deal with particular considerations that might vary among fisheries. It is neither constructive nor facilitative of research for FCB to propose changes which reduce rather than expand the flexibility of the original model, which was itself an extension of, not a replacement for, the classical literature. Neither that literature, nor my papers, are so rigidly defendable as "truth."

* University of Massachusetts. I am grateful to the National Science Foundation for research support.

distribution may be an important determinant of cost (as well as fish reproduction). I hope it is not self deception for me to think that I have made myself clear!

II

So much for the general treatment of cost, which is of course essential to a general theory of natural resource production since different resources, even different fisheries may exhibit very different cost behavior. In the case of fisheries, FCB propose that very restrictive assumptions be imposed on the cost function. However, the reader should be warned that they are wrong in thinking that my model and my analysis does not fully include the situation they want to represent.

The case treated by FCB is precisely my model with a cost function given by

$$(1) \qquad C(x, X, K) \equiv \begin{cases} \hat{\pi}, & x \leq g(X, K) \\ \infty, & x > g(X, K) \end{cases}$$

The theory of the firm says that output is expanded as long as price is not below marginal cost. Since FCB are compelled by their assumptions to believe that marginal cost is zero up to the output given by $g(X, K)$, and infinite for greater outputs (i.e., larger outputs are impossible) this means that the firm's profit maximizing output is $x=g(X, K)$ which is consistent with my model, taking account of the corner conditions of maximization.

III

FCB's assumptions about cost are thus not precluded by my model, but the reader may wonder, as do I, why their faith in such a severely restricted cost function is held with such conviction.[2] They would have us believe, for example, that no matter how large the stock of fish, the firm continues to choose the same input combinations and incur the same total cost: that the fisherman will have no incentive to change his gear (size or number), to change his intensity of use of that gear, fuel consumption, etc. Their imagination is too limited to admit the possibility that a doubling of the stock of fish would not only increase output because the net or trap hauls are larger (a "windfall"), but also because less time is spent finding or chasing the fish or hauling in half-empty gear, so that an extra run to the fishery may be accommodated. It would be surprising if output, fuel, labor, and gear costs were not affected by all these considerations, but this is really an empirical matter on which FCB offer only dogmatic assertions that output and total cost cannot be varied by the firm. I hope that fisheries policy does not follow farm policy where it was once thought that removal of land from cultivation would be adequate to reduce agricultural surpluses, as if the intensity of fertilization and other aspects of cultivation were invariable.

IV

As noted above, the question of which special cost function is most appropriate is an empirical question. For empirical work I would suggest[3] a marginal cost function of the form indicated by curve A in Figure 1, rather than B as proposed by FCB. The hypothesis would be that the firm faces an upper bound on its catch rate given by $\bar{x}=g(X, K)$. No matter how intensively the firm fishes, it could not exceed a certain catch rate determined by the biomass and number of firms. But it can approach $\bar{x}$ more or less closely by varying inputs and therefore cost with marginal cost becoming steeper the nearer x is to $\bar{x}$. For example, the function

$$C' = 1/(\bar{x} - x)^{\alpha}, \; x < \bar{x}, \alpha \geq 1$$

has this property. Furthermore, for $x<\bar{x}$, $C' \rightarrow 0$ as $\alpha \rightarrow \infty$ so that marginal cost B is a limiting case of A. If empirically one obtained very large estimates of α from firm

---

[2] A simple industry model like that of FCB has already been presented in James Quirk and V. L. Smith, (pp. 29–30), but I trust no one will view it as the last word in fisheries theory, instead of an illustration using restricted assumptions.

[3] I hope no one will take me too literally when I use the word "suggest." I would not recommend that serious empirical work begin until the fishery harvest process is stochastically modeled, since it is obvious to anyone who has ever wet a line, that an important feature of commercial fishing is the uncertainty of the catch. The firm's inputs, including (X, K) are most accurately thought of as parameters in a probability distribution function defined on harvest quantities.

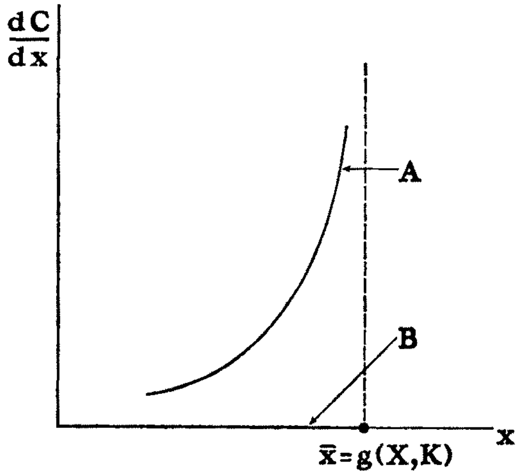


FIGURE 1

cost data, this would lend support to the FCB faith, otherwise not. What is crucial for empirical work is a cost hypothesis which does not exclude the possibility that cost will depend on x, X, K, and perhaps other variables, and which is rich enough to include special limiting cases.

V

I leave it to the reader to puzzle out the meaning of the FCB rhetoric in the four sentences immediately following equation (2).

VI

Equations (3)–(7) are described by FCB misleadingly as the "classical system," but, as the reader can verify, it is just a reproduction of my model using a special "classical" cost function (1). In fact they even employ the same phase diagram representation in Figures 1–4. The traditional literature, cited by FCB, does not employ these tools and does not state a dynamic system such as (3)–(7) modeling the bionomic interaction of industry with the resource. It is reassuring that except for the restriction to a particular form of the cost function, the rest of my original model is embraced warmly by FCB.

VII

FCB claim that my figures illustrating the set of points $F(X, K)=0$ corresponding to biomass equilibrium are "erroneous" in showing a positive slope, and that the "error is directly traceable to the misspecification of the cost function." Both statements are false.

As to the first statement, I showed on page 418 (also in the Appendix, page 430) that the sign $dK/dX|\dot{X}=0$ can be positive or negative depending on the signs of derivatives of the cost, revenue, *and* fish growth function, f. I further showed that if $C_{13}=0$, $C_{12}<0$, and $C_{11}>0$, then the sign is either (i) always negative or (ii) nonegative for all $X \leq \hat{X}$ (where $\hat{X}$ is a point such that the slope is zero). Surely it must be obvious from this that I was well aware that my model is fully consistent with $F(X, K)=0$ defining an inverse relationship for all (X, K). However, because the model showed the possibility of a positive slope, Figures 3–5 were drawn up so that "in each case it is assumed that (the sign) is positive below some value of X, and negative above that value." It would have been erroneous to do otherwise.

The second statement is wrong because their special assumptions about cost are neither necessary nor sufficient for the set $F(X, K)=0$ to have a negative slope. This can be proved by noting that their equations (3) and (4) define $\dot{X}=F(X, K)=f[X, Kg(X, K)]=0$. Hence,

$$\left.\frac{dK}{dX}\right|_{\dot{X}=0} = -\frac{f_1+f_2Kg_1}{f_2(Kg_2+g)} \gtrless 0,$$

where $g_1>0$, $g_2<0$, $f_2<0$, $f_1 \gtrless 0$.

FCB get the negative sign throughout by placing still more restrictions on their model (3)–(7). They use a homogeneous quadratic for $f(X)$, and set $g(X, K)\equiv rX$, with the result that $dK/dX|_{\dot{X}=0}$ is negative and constant. But this negative slope can be obtained with much less restrictive assumptions. Here is an example: Let $\dot{X}=aX-bX^2-Kx$, and $C=(\gamma x^2/X)+\hat{\pi}$. Then setting price equal to marginal cost gives $p=2\gamma x/X$, $\dot{X}=F(X, K)\equiv aX-bX^2-(p/2\gamma)KX$, and therefore when $\dot{X}=0$, we have $K=(2\gamma/p)(a-bX)$. The implication is clear. If the set $F(X, K)=0$ has a negative slope in the phase space, this implies *absolutely nothing* about any particular form for the cost function. What a negative slope does is to put a condition on the derivatives of both C and f. If price is constant, the condition is (p. 418)

$$\left.\frac{dK}{dX}\right|_{\dot{X}=0} = \frac{KC_{12} + C_{11}f'}{xC_{11} - KC_{13}} < 0,$$

which should be the starting point for FCB if they want to explore the implications of an inverse relation between X and K.[4]

VIII

FCB are wrong in thinking that when $f' > 0$, the policy implication is that the industry should be subsidized. The statement $\lambda = KC_2/f'$ does not hold (this is just elementary constrained maximum theory) *except* at a stationary state maximum point for the firm under appropriation, and $f' < 0$ at a maximum. It follows that $\lambda > 0$. When FCB try to interpret the tax as a subsidy "when" $f' > 0$, they do so incorrectly on their own responsibility. I made no such interpretation, and I very much hope the fishing industry will never suffer from such misinterpretations.

Certainly under free entry we might have equilibrium with $f' > 0$, but this is also not socially optimal and by no stretch of the imagination can one state that $\lambda < 0$, calling for a subsidy. Starting from such an initial position, if the socially optimal equilibrium stock is obtained, say by placing a temporary moratorium on fishing, that stationary equilibrium could then be maintained by permitting free entry under the social charge, λ.

Now the question of how you get an industry to a maximum (the optimal time path), as distinct from how you keep it there after it has arrived, cannot, and was not intended, to be answered by a stationary state model. For this purpose one needs a model of maximization over time (see Quirk and Smith, and Oscar Burt and Ronald Cummings). In such a model the appropriate social charge will vary over time as the capital and fish stocks change, until a stationary equilibrium is obtained.

IX

It is not illuminating to say that "the classical theory can define a tax which depends upon the parameters but not the slope of the biological growth function." *Any* theory can do this if its assumptions are regarded as variables to be adjusted until that result is obtained! FCB get this result by treating the special case discussed above in Section VII.

X

The dynamic paths shown by FCB in their Figure 1 are not correct, i.e., such paths do not follow from their differential equation system. Refer to the derivation of the path conditions (A.3) and (A.4), and the phase diagrams in my paper.

REFERENCES

O. R. Burt and R. G. Cummings, "Production and Investment in Natural Resource Industries," *Amer. Econ. Rev.*, Sept. 1970, *60*, 576–90.

J. P. Quirk and V. L. Smith, "Dynamic Economic Models of Fishing," *in* A. Scott, ed., *Economics of Fisheries Management: A Symposium*, Vancouver 1970, pp. 3–32.

V. L. Smith, "The Economics of Production from Natural Resources," *Amer. Econ. Rev.*, June 1968, *58*, 409–31.

---

[4] A fairly general set of sufficient conditions for a negative slope are the following: no crowding externalities ($C_{13} = 0$); $f(0) = 0$; and a cost function homogeneous of degree one. The cost function can then be written $C(x, X) = X\psi(x/X)$. Hence, $C_{11} = \psi''/X$, $C_{12} = -(x/X^2)\psi''$, and the above expression reduces to

$$\left.\frac{dK}{dX}\right|_{\dot{X}=0} = \left(\frac{1}{x}\right)\left[f' - \frac{f}{X}\right] < 0,$$

since $f(0) = 0$, $f'(0) > 0$, and $f'' < 0$ imply that $f' < f/X$, for $X > 0$. That is, marginal growth in the fish stock is always below average growth. But $f(0) = 0$ rules out the possibility of a critical species population below which it is not viable. As to cost homogeneity, I am skeptical because fishing units are confined to operate on a plane while many prey occupy a volume below. This strongly suggests a surface volume relationship underlying the physics of mass contact between predator gear and prey. Cost would then not be homogeneous since for given fish and vessel densities, the fish mass will vary roughly as the cube while fishing units will vary as the square of the radius (or other lineal measure) of the contact region.

Finally, bear in mind that all of this assumes price remains constant in the face of changes in the total fishery catch. If we drop this rather unrealistic assumpthen, then the above conditions are not sufficient to yield unambiguously a negative sign for dK/dX (see p. 430 of my original paper for the general case where price varies with the harvest).

Employment and Rural Wages in Egypt: A Reinterpretation

By JAMES A. HANSON*

Within the last two years a series of articles has been published dealing with the existence of surplus labor in the United Arab Republic. The most recent, an article in this *Review* by Bent Hansen, used the results of a rural employment survey conducted by the International Labor Organization and the United Arab Republic government (*Employment Problems in Rural Areas*, henceforth *EPRA*) to "settle the issue of surplus labor rather definitely against the surplus labor hypothesis" in that country (Hansen, p. 298). In the present author's view, neither Hansen's article nor the *EPRA* survey itself provide the evidence for such an unqualified rejection of the surplus labor-traditional wage hypothesis; nor do they suggest unqualified acceptance of a marginal productivity theory of distribution. Instead, it will be argued that the evidence is broadly consistent with more sophisticated hypotheses about rural wage determination. This latter view holds that with correct policy, labor might be drawn out of the agricultural sector for work on development projects or in regional industry, at relatively constant wages.

I. The Theory of Wage Determination in Peasant Economies with Abundant Labor

The main thrust of Hansen's article is contained in his statements "that farmers have a real choice between taking paid labor outside the farm or working on the farm. Under these circumstances there is little reason to believe that the value of the marginal product of labor, even on the smallest farms, would be smaller than current rural wages" (p. 302). This is particularly true for male adults who "seem by and large to be fully employed with long working hours during spring and summer and some underemployment from October to February" (p. 311).

Leaving to later sections the empirical question of whether paid, nonfarm employment is available, Hansen's theoretical argument that the marginal product of labor is positive and equals the wage is not strictly correct. Although the surplus labor theorists have never satisfactorily resolved the apparent contradiction between the existence of positive wages and zero marginal products,[1] it must also be admitted that the mere payment of wages does not imply competitive wages which are equal to the marginal product of labor. Particularly in the context of an underdeveloped, peasant, agricultural economy, a variety of noncompetitive distribution systems which involve wages are possible.

For example, assume that the rural economy largely consists of small farms, operated mainly for subsistence by families of homogeneous laborers. Next, assume that a laborer must be physically present to obtain his share of the implicit land rents earned by the family plot and that the offers of paid employment are some distance from the family plot. In this case, laborers would respond to the difference between the market wage and the average product of labor on the family farm, not the marginal product. Wages would, therefore, exceed the marginal product of labor on the family farm. The marginal product is not relevant because the economy has not developed institutions that permit the collection of property income without the need for physical presence. Such a model of a rural economy certainly has some applicability to Egypt, for the bulk of hired labor comes from small farms operated by large families (*EPRA*, II, Part II, Tables 37–60), and various observers have noted the long distances which workers need to travel to obtain employment (*EPRA*, I, pp. 54–55; Mead, pp. 96, 99).

---

* Brown University.

[1] For some attempts at resolving this problem, see John Fei and Alpha Chiang, and Harvey Liebenstein. Leibenstein's argument that hired labor must work harder and be paid more has been applied to Egypt by Donald Mead.

Alternatively, one could assume that holding land and hiring labor confer social status, while working for wages reduces it. According to Mead and H. Habib Aryout, two commentators on rural Egypt, these views are common among Egyptian peasants. Under this assumption, workers could be obtained only by paying wages in excess of the marginal product of labor on subsistence farms; while employers would be willing to pay more than the marginal product of labor on their farms to gain the "consumption aspects" as well as productive benefits of labor. The social preference for self-employment would lead to a wage rate which exceeded the marginal product of labor on subsistence farms. Furthermore, the social preferences might manifest themselves in other ways: even small farms might hire extra labor for "consumption purposes," very small farms might consist of separate, unconsolidated plots in order to prevent the owner's neighbors from learning of the small size of his holdings and his need to work for pay, and peasants might travel considerable distances to seek employment, rather than face the degrading need to work for their own peers. All of these corollaries of the positive preference for self-employment are prevalent in Egypt.[2]

If we ignore the possibility or importance of "market imperfections," highlighted by the above arguments, Hansen's argument is correct, but it may be irrelevant. Although the peasants may actually equate on-farm marginal products with off-farm earnings, there may be few job offers during the slack season, October to February. Thus the average farmer would be forced to occupy himself with low productivity, nonagricultural pursuits during this time. Correspondingly, employment opportunities for women and children who engage in these activities in the busy season would be greatly reduced. This argument has been hinted at by Mead and by Hansen himself (pp. 304, 311). The remainder of this note is devoted to providing evidence which supports this view and, incidentally, casts doubt on Hansen's first point that wage labor is a viable alternative to on-farm work.

II. The Extent of Unemployment and its Seasonality

Hansen's argument that surplus labor is nonexistent and the labor force is fully employed is based on a special interview subsurvey taken during one week in January, at the depth of the seasonal trough (Hansen, pp. 304–06; *EPRA*, I, pp. 157–61). However, the quoted unemployment rate of 8 or 9 percent during this slack period is actually lower than the 12.5 percent average *annual* rate among adult males in agricultural enterprises which is cited in the complete survey (*EPRA*, I, p. 183). Since the methodology that was used to calculate these two conflicting figures leaves something to be desired,[3] we have turned to the basic data re-

---

[2] See Hansen, (p. 303); *EPRA*, I, (pp. 54–56, 67–70); II, Part II, Tables 37–67; Mead (p. 97).

[3] The figures cited by Hansen come from a special labor survey taken during one week in January when unemployment is highest (*EPRA*, I, pp. 157–61). According to their responses potential workers were divided into jobless (52 percent) and working (48 percent). Of the jobless, 81 percent were unemployed for reasons of school attendance, old age, physical handicaps, or military service. One percent were employers who were regarded as working, leaving 18 percent of whom 4 percent were workless because of lack of opportunity, 9 percent because of tradition, and 5 percent who did not reply. The division of these figures into an unemployment rate of 8 or 9 percent of the labor force is rather hazy at best. Nor does the rate include underemployment of those persons actually working during the reference week. These factors might easily yield figures of well over 10 percent unemployment, which correspond to those obtained from the full survey's hourly data described in the text.

As to the figure of 12.5 percent unemployment contained in *Report C* of *EPRA*, I, there is no accurate description of which subset of the full survey was used to obtain the figures for hours worked by provinces and regions, nor the methodology by which it was selected (*EPRA*, I, pp. 214–23; based on *EPRA*, II, Part II, Tables 1–18). Using the methodology of *Report C* to estimate unemployment (303 eight-hour days), no reasonable subset of labor input by strata taken from the full survey (*EPRA*, I, pp. 224–228; II, Part II Tables 37–65) gives unemployment rates that approach the 12.5 percent figure. Finally, when the regional figures ported in *EPRA* and calculated our own annual unemployment rates.

According to the survey tables (*EPRA*, I, pp. 100–02; II, Part I, pp. 58–67), men[4] were completely unemployed an average of 70 days while men in agricultural households were idle 81 days. Assuming, following the survey, that workers rest on Fridays and holidays, there are 303 possible working days (*EPRA*, I, pp. 178–79) and the resulting unemployment rates are 5.6 and 6.3 percent, respectively. Turning to the hourly figures, on 28 of his working days the average man worked less than 6 hours, while the average man in an agricultural household worked less than 6 hours on 24 days[5] (*EPRA*, I, p. 101; I, Part I, pp. 62–63). However, on other days they were often employed for more than 8 hours, and depending on which subset of the survey is chosen, the average workday ranges from 8.5 to 8 hours. Therefore, if we again follow the survey and assume that an 8-hour day is the norm,[6] the rates of unemployment based on hourly data are about the same as, or sometimes much less than those based on days worked. As Hansen points out, these results are only slightly lower than the unemployment percentages obtained through the interview survey during the slack season and thus support his hypothesis of full employment.

However, in any estimate of the stock of unemployed labor the norm must be chosen with care and in this case the norms seem much too low. First, it is likely that many more than 303 days are available. Another special subsurvey, taken in April, shows no difference between the number of hours worked on Fridays and other days. Moreover, the major religious holiday, Courban Bairam, was observed with only 1 day of rest, rather than the 4 assumed in the survey. In the words of the survey: "In practice, they (the farmers) work if there is work available and they rest if there is a lack of work."[7] Correcting only for the understatement resulting from the holidays of Courban Bairam and Ramaden Bairam would add roughly 2 percentage points to the unemployment rate.

The survey's norm of 8 hours worked in each of these days is also too low and seems to be based on Western industrial practices rather than non-Western peasant agriculture.[8] Its inapplicability becomes particularly obvious when we turn to a discussion of the seasonality and regional factors in unemployment. Although there is a variation depending on the type of crop, broadly speaking, there is a busy season with intermittent lulls from March through September with a slack season after the last crop, cot-

---

that were used to calculate the 12.5 percent figure are divided between Upper Egypt and the Delta, there is no significant difference in employment rates. All other evidence contradicts this view. Thus it seems best to use the full labor survey rather than this undocumented subset.

[4] In this case men refers to adult males who are permanent members of a household. Workers are regarded as permanent if they spend over thirty days working for the same farm. Although no temporary workers per se are included in survey, there are two strata of landless farmers. Since their unemployment rates are higher, to the extent that the survey includes too few of this group the unemployment figures will be too low. Figures for men are used here because the availability of children and the division between household and farm work are uncertain. Moreover, in a traditional, nonmatriarchal society unemployment among men should indicate even greater unemployment among women and children. However, these classes are not necessarily reservoirs of labor for, as income rises and educational opportunities increase, these groups may withdraw from the labor force. Counteracting this tendency is the decline in tradition mentioned by Hansen and the shift from production of household goods, or subsistence handicrafts to specialization and purchase in the market with a decline in the time necessary to maintain the house. The causes and effects of this shift have been described recently by Stephen Hymer and Stephen Resnik.

[5] Figures for average number of days worked and average annual hours worked refer to slightly different groups of laborers.

[6] The survey claims that it used a 7-hour day (*EPRA*, I, p. 179) but the actual figures for weekly and annual available labor are divisible by multiples of 8 and not 7 hours (*EPRA*, I, pp. 212–28).

[7] See *EPRA*, I, (pp. 107–109, 182). If we assume, following the survey, that there are only 3 working days of 8 hours in the survey period containing Courban Bairam and 4 working days of 8 hours in the period containing Ramadan Bairam, the overemployment is 213 and 129 percent for all workers. The corresponding figures for agricultural workers are 187 and 116 percent (*EPRA*, I, pp. 182, 214, 217).

[8] On the problems of applying Western norms see Gunnar Myrdal's *Asian Drama*, ch. 21, 22, and Appendix 6.

ton, is harvested. During this peak season, the average number of hours worked increases significantly, from 6.4 to 8.8 per day.[9] Assuming that enough calories are available to make it physically possible to work at the peak rate throughout the year, this fact implies that a substantial number of hours are available in the present slack period for work on a second crop, infrastructure, or in cottage industry; hours which could be worked without interfering with the present production of food or cash crops. In fact, on this basis, and correcting for the errors in the holiday weeks described above, annual unemployment among permanent male workers in agricultural enterprises amounts to 10 or 11 percent, almost double the uncorrected survey figure of 5 to 6 percent.

The large seasonal variation in employment also casts doubt on Hansen's contention that workers have a real choice between taking paid labor and working on the farm. In fact, although paid employment outside the household enterprise represents 20 to 25 percent of the men's work time, 65 percent of this work occurs during the busy season (*EPRA*, I, p. 69; II, Part II, p. 68, Tables 18.a, and 37–60).

Of course, one could argue that workers are idle in the off-season by choice rather than through the lack of employment offers. In that case, there is no surplus labor, for potential employers would have to overcome the workers' preferences for leisure through higher wages and few additional workers could be obtained at the going wage rate. To settle this point a knowledge of the response to wages, which is discussed in the next section, and, ultimately, preferences for income and leisure is necessary. However, there is some indirect evidence that measured unemployment is due to lack of demand for labor rather than supply factors.

According to the survey a great disparity exists between employment rates in Upper Egypt, which has only one crop a year in most areas, and the Delta, where irrigation and two crops are more prevalent (Hansen, p. 303; *EPRA*, II, Part II, Tables 10A, 12A, B). Assuming regional attitudes toward work are similar, this suggests that much of the unemployment is technological (and regional).[10] Without the irrigation which makes a second crop possible, there is little to do in Upper Egypt during the off-season. Throughout the country a large percentage of working time during that period is spent on nonagricultural and nonfarm jobs such as animal watching, marketing, and production of handicrafts (*ERPA*, I, p. 151 and Graph 20) that have an obviously low productivity. It seems unlikely that much time would be spent in this way if, as Hansen contends, wage employment existed. If more remunerative or productive jobs were made available in the off-season, this low level work might be shifted to women or children, reducing their unemployment. As Hymer and Resnik have pointed out elsewhere, this shift to more productive work would be even more likely and wages would rise even less, if the output of the new jobs was a good substitute for the production of subsistence handicrafts.

---

[9] See *EPRA*, I, (pp. 110–54). A statistical test of this difference is easily made. Assume that the total weekly hours worked by those sampled in each of the strata (*EPRA*, II, Part II, Tables 37–65; I, pp. 212–28) are a random variable, influenced by changes in weather, etc. (Owing to the problem of Fridays discussed in the text, it was felt that total weekly hours was the relevant variable. However, to maintain consistency with the previous paragraphs, figures in the text were calculated by excluding holiday weeks and assuming no work was done on Fridays.) Under this assumption the difference between the average hours worked per 8-day period in the busy and slack season is statistically significant at the 5 percent level, for each of the 8 strata of permanent male agricultural labor. The strata, listed in Hansen's article, are the result of cross classification by farm area and family size.

[10] Mead's calculations support this result. As Hansen points out (pp. 303–04, 311), there is an even greater regional disparity in unemployment rates among women and children, so technological unemployment is underestimated by this calculation. Of course, these disparities may be due to regional differences in traditions and attitudes toward work and leisure, but traditions change, and, if the change occurs through greater offers of work at the going wage we have a surplus labor situation. One discordant note does arise over the use of this data which is from the full survey but based on labor inputs. The regional employment data contained in the full survey and described in fn. 3 do not show a similar variation between regions. As discussed in that footnote the regional data is judged to be unreliable, but if the reader decides it is preferable, he also must accept the resulting unemployment rate, 12.5 to 15 percent.

Another test of the hypothesis that unemployment is due to choice rather than technology is a comparison of hours worked per permanent worker on different size farms; if employment were readily available then there should be no difference in the time worked per permanent worker. However, the results are dominated by farm size; in all cases save one, the larger the farm, given the family size, the more total hours worked by the average permanent male worker. Also, seasonal fluctuations in employment are much larger on the smaller farms, mainly as a result of the seasonality of outside work (*EPRA*, I, pp. 118–120). These results seem to indicate that additional work on small farms runs into diminishing returns, and contrary to Hansen, outside work is not always available.[11]

III. Rural Wages

As further support for the hypothesis of marginal productivity distribution, Hansen presents wage data from the survey (*EPRA*, III). Both the method of compiling the data—asking the village sheik for prevailing wages—and the procedure of excluding villages in which a substantial number of observations were missing would seem to bias the result toward the traditional wage theory. However, according to Hansen, the distribution of the wage data is not truncated, or squeezed up against a traditional subsistence minimum, but fairly normal, disproving the traditional subsistence wage theory once again.

Of course, as shown earlier, disproving the traditional subsistence wage theory is not the same as proving that distribution is by marginal products. Moreover, Hansen's argument against the traditional wage theory is faulty on at least 3 counts. First, out of the original sample of 48 villages, 17 were eliminated as having missing information, particularly during the slack season (Hansen, p. 306; *EPRA*, III). This fact casts still further doubt on Hansen's contention that nonfarm wage labor is a viable alternative to farm work throughout the year. Instead it supports the view, expressed earlier, that there is large seasonal and, since most of the missing villages are from Upper Egypt, regional technological unemployment. Second, only the most naive proponent of the subsistence wage theory would argue that workers react to money wages; the argument is usually phrased in terms of the money cost of a subsistence bundle of goods. However, the wage data presented by Hansen are undeflated by regional prices. The substantial variation in average money wage between govornerates, particularly between Fayoum, an oasis in Upper Egypt which is separated from the Nile Valley by 10 miles, and the rest, is certainly evidence of some imperfect connections between factor and/or goods markets. Third, if we decide to use the data in spite of this obvious deficiency, but omit the 6 villages in Fayoum, which are the only observations from Upper Egypt, the distribution appears truncated rather than normal. Statistically speaking, neither the hypothesis that the data are normally distributed, nor the hypothesis that they are distributed as a truncated half of the normal distribution can be rejected at the 5 percent level.[12] Thus this evidence cannot be used to reject either the marginal productivity theory or the traditional wage theory of distribution.

---

[11] An analysis of variance test of average hours worked per man, based on a table classifying farms by area and family size, shows the differences in labor employed in different size farms are significant at the 10 percent level. The data are the calculated average annual hours worked by permanent male employees in each strata and were taken from the survey (*EPRA*, II, Part II, Tables 37–65 and *EPRA*, I, pp. 212–28). These results also show up in Hansen's labor input data (Table 2, p. 301) and were mentioned in the survey (*EPRA*, I, pp. 118–20). Again the result might be due to a difference in tastes, since some farmers may have been able to purchase larger farms because they were harder workers and accumulated more wealth.

[12] A Kolmogorov-Smirnov test of the goodness of fit to a hypothesized distribution, as described in B. Lindgren and G. McElrath, was performed on the observations from the remaining 25 Delta villages (Hansen, Table 5, p. 297). Under the null hypothesis that the sample was drawn from a normal distribution with mean and standard deviation of 18.79 and 2.71 piasters per day, respectively, the largest difference between the cumulative frequency of the sample and the hypothesized normal distribution is .20. This is a smaller difference than we would get under the null hypothesis 95 percent of the time and so we cannot reject the null hypothesis. For the truncated distribution the observed

Even more puzzling is Hansen's discussion of variations in wages. He states that there is substantial seasonal fluctuations in daily wages, since the highest daily wage for men is 150 percent of the lowest, and he uses this variance to support his claim of great flexibility in wage determination. However, no other figures are presented on the variation. Moreover, part of this variation could obviously be accounted for by the inflationary trend in the case of the subsistence market basket and part by the variation in the average number of hours worked per day.[13] In fact, Hansen's regression equation of average daily wages on hours and a time trend is

$$W_t = 5.630 + \underset{(.023)}{.219H} + \underset{(.008)}{.073t} \qquad R = .88$$

where

W_t = average daily wages of men at time t
H_t = average hours of work per week (all men)
t = time t = (0, 1, . . . , 52)

and all coefficients are highly significant. Hansen interprets the regression as supporting the hypothesis of demand and supply determined wages by identifying average hours of work per week as a proxy for days worked and assuming that demand shifts while the supply curve has a positive slope and is fixed.

It should be clear however, that there is another interpretation of the regression equation which strongly supports the traditional wage-surplus labor view. The alternative hypothesis, consistent with the labor surplus theory of wage determination and supported by the regression, is a horizontal or nearly horizontal supply curve of labor (measured in hours) at some constant real hourly wage.[14] Neglecting the constant, for which no standard error has been provided, and assuming the time trend represents only inflation, the regression coefficient for weekly hours, divided by a constant equal to the available man-days per week, is equal to the constant, average, real, hourly wage.[15] Under this hypothesis average daily wages would then be determined by shifts in the demand for labor hours against a horizontal supply curve of labor, also measured in hours.

The hypothesis of a horizontal or nearly horizontal supply curve is also supported by some alternative calculations. The annual difference between hourly wages in the two overlapping weeks (June 21 and 28, 1964 and 1965) in the wage survey is 37.5 percent and 37.3 percent, respectively (*EPRA*, III, p.

---

frequency table of daily wages was inverted between the second and third classes (containing 2 and 7 observations), and the resulting set of new and sample observations were summed by classes. This procedure results in a normal distribution which is symmetric about a mean of 15.95 piasters per day, with standard deviation 3.92. Under the null hypothesis that the sample was drawn from the right tail of this distribution, we find that the largest difference between the cumulative frequency of the sample and that expected under the null hypothesis is also .20. Again we cannot reject the null hypothesis at the 5 percent level of significance. In other words, the available data does not permit the investigator to distinguish between the two hypotheses.

[13] Hansen assumes laborers are hired on a daily basis and there is little seasonal fluctuations in hours worked. He admits this point is crucial to his interpretation of the regression (p. 309). However, as shown earlier, the seasonal fluctuation in daily hours worked is significant.

[14] There is no reason to think that Hansen's daily wages rather than hourly wages are the correct price variable. One could easily imagine a traditional hourly wage. If leisure has any value and workers can vary their hours, hourly wages are the correct variable. As R. Albert Berry and Ronald Soligo have shown, except in the case of leisure satiation this implies a rising supply curve of labor, although the exact elasticity is not discussed. Our argument in this paper implies that the elasticity of labor supply is very high. Another possibility which would result in a horizontal labor supply curve is constant labor productivity in the production of subsistence handicrafts, which, according to Hymer and Resnik, are the real alternative to field work.

[15] If the constant is statistically significant it might simply be a statistical reflection of an accounting or traditional tendency to hire laborers for some fixed number of hours per day. This proposition could be tested either by regressing daily wages against a constant and the average number of hours worked in execss of the fixed number, and testing for the equality of hourly wages for regular time and overtime, or most simply by regressing hourly wages against a constant, hours, and the time trend. A rough version of the second test was performed in the text.

37). Ignoring growth in the labor force, and making Hansen's assumptions about the equality of time worked during the same periods in the year, this figure represents the inflationary trend.[16] A rough series of real hourly wages can then be obtained by deflating each of the 54 weekly average hourly wage figures in the 31 village sample by the inflationary trend. Although the real hourly wage is 10 to 15 percent higher in the busy season—March to September—if we treat the weekly observation as a random variable, this difference is not significant at the 5 percent level. Thus we accept the hypothesis of a constant, traditional hourly wage. On the other hand, as discussed earlier, there is a statistically significant seasonal variation of 25 to 30 percent in weekly hours worked by males in each agricultural strata. This supports the view of a shifting demand curve for labor intersecting a relatively horizontal supply curve.

IV. Summary and Policy Implications of the Alternative Hypothesis

This note has argued that the evidence presented in Hansen's article and the *EPRA* survey on which it was based is broadly consistent with either a marginal productivity theory of distribution, or a traditional subsistence wage theory of distribution. However, the evidence is most consistent with a more sophisticated hypothesis of wage determination incorporating both substantial seasonal variation in employment and handicraft production.

What are the policy implications of accepting the more sophisticated hypothesis rather than the marginal productivity hypothesis of wage determination For example, what do the different hypotheses imply about the impact of the Aswan Dam? Should the Egyptian development effort be biased towards large scale manufacturing, smaller scale, regional dispersed manufacturing, or agricultural improvements?

To answer the first question, the more sophisticated view, encompassing the seasonality of productive labor and the variations in employment by regions and size of holdings, seems to indicate the possibility of really significant increases in food output through the greater use of irrigation, which would make two-crop rotation possible in Upper Egypt, the development of crops which mature in shorter seasons, and the use of fertilizers to prevent declining soil fertility. At present, five weeks are lost in January owing to the lack of water. The Aswan Dam will make this loss unnecessary (*EPRA*, I, pp. 117–18) and if crops can be modified to mature in a shorter time, three crops may be possible, provided increased amounts of fertilizer are used to prevent declines in soil fertility. If workers were presently fully employed throughout the year in productive labor, as implied by Hansen's marginal productivity theory, such output increases would be much less.

To answer the second question, there seems to be some possibility of building infrastructure, transportation networks and then developing regional industry using off-season labor or the present landless peasants; this would be particularly true if the output of these projects yielded substitutes for the present uses of off-season labor, e.g., more durable irrigation works and farm machinery. In this way, output and capital would be increased without a decline in food production. Based on the combined results of the wage and employment survey described in Sections II and III, rather than a marginal productivity theory of income distribution, wages would not have to be increased much above prevailing levels to obtain the necessary labor for these projects.

REFERENCES

R. A. Berry and R. Soligo, "Rural Urban Migration, Agricultural Output, and the Supply Price of Labor in a Labor Surplus Economy," *Oxford Econ. Pap.*, July 1968, *20*, 230–49.

J. Fei and A. Chiang, "Maximum Speed Development through Austerity," in I. Adelman and E. Thorbecke, eds., *The Theory and Design of Development*, Baltimore 1966.

---

[16] Hansen's figure, calculated from the regression, is about 30 percent.

H. Habib Aryout, *The Egyptian Peasant,* Boston 1968.

B. Hansen, "Employment and Wages in Rural Egypt," *Amer. Econ. Rev.*, June 1969, *59,* 298–313.

S. Hymer and S. Resnik, "A Model of an Agrarian Economy with Non-Agricultural Activity," *Amer. Econ. Rev.*, Sept. 1969, *59,* 493–506.

C. Kao, K. Aschel, and C. Eichner, "Disguised Unemployment in Agriculture: A Survey," in C. Eichner and L. Witt, eds., *Agriculture in Economic Development,* New York 1969.

H. Leibenstein, *Economic Backwardness and Economic Growth,* New York 1963.

B. Lindgren and G. McElrath, *Introduction to Probability and Statistics,* New York 1959.

D. Mead, *Growth and Structural Change in the Egyptian Economy,* Homewood 1967, pp. 90–98.

G. Myrdal, *Asian Drama,* New York 1968.

Institute of National Planning, *Employment Problems in Rural Areas,* U.A.R.

(I) *Report C on Utilization of Rural Manpower and Measurement of its Underemployment,* prepared by M. Mohiey El-Din Nazart, S. Shabana, S. Rofael, U. Planck, and Q. Qayoum, Cairo, Aug. 1966.

(II) *Statistical Tables, The Labour Record Sample Survey,* Part I, Cairo, Dec. 1965; Part II, Cairo, Sept. 1966.

(III) *Report D on Wages, Income, and Consumption in Rural Areas,* prepared by B. Hansen, A. Sedki, and Y. Moustafa, Cairo, Dec. 1965.

Employment and Rural Wages in Egypt: Reply

By Bent Hansen*

James Hanson has reinterpreted the data on employment and rural wages in Egypt which were collected by the rural employment survey (*Employment Problems in Rural Areas* (*EPRA*)) of the International Labor Organization and the Institute of National Planning, Cairo. Moreover, he has presented some qualifications to my interpretations which are justified, but his basic alternative interpretation of the seasonal fluctuations of daily wages does not hold water. This alternative interpretation was, in fact, carefully examined and rather effectively rejected by the *EPRA* report itself on the basis of information which Hanson, unfortunately, has overlooked.

Before I enter upon details, I want to make it clear—Hanson's exposition of my views gives the opposite impression—that there is no disagreement about the existence of some *seasonal* unemployment and underemployment, and the possibility that there may be slack periods during the year with temporary zero marginal productivity of labor in agriculture. I summarized my article (1969, p. 311) by saying that "*Male adults* seem, by and large, to be fully employed with long working hours during spring and summer, and with some underemployment from October to February"; that for the year as a whole "*Female adults* . . . are slightly underemployed . . . "; . . . "children between six and fifteen work, on the average, slightly more than women outside the households." I found it rather meaningless, although, to apply the notion of underemployment to children between six and fifteen (if anything, I found them overemployed). Thus I did not argue, as Hanson maintains, that surplus labor is nonexistent and the labor force is fully employed. When I said (p. 302), that "Farmers (i.e., men), by and large, are fully occupied, taking into account the work outside their own farms," this was a statement which referred to total hours worked per year without considering the distribution during the year. Nowhere in my article did I maintain that work outside the farm is a viable alternative to farm work throughout the year i.e., at any time of the year. Hanson's efforts to prove that men are to some extent unemployed at certain times of the year and that outside work is not always available only repeat what I thought was obvious from my article.

As I see it, the major issue in the rural surplus labor problem is: does there exist a surplus of labor in the sense that marginal productivity of labor at all times of the year is zero (or at least, very low) so that labor could be withdrawn permanently from agriculture without detrimental effects on output; or, is the surplus of a purely seasonal nature and if so, how large is the seasonal surplus and how long a time of the year does it exist. It is the first case which traditionally has been considered in development theory and which makes demand-supply theory of wages break down so that if wages actually are paid, some sort of "institutional" wage explanation is called for.[1] The existence of seasonal slacks and seasonal underemployment gives rise to entirely different development problems and is fully compatible with demand-supply theory of wages even in a fully competitive labor market and even if marginal productivity occasionally during the year should become zero. I shall discuss this more in detail below. The *EPRA* shows unequivocally that surplus labor in Egypt is a seasonal problem and of much less importance than has hitherto been assumed. Hanson has made no attempt to deny that.

---

* Professor, University of California, Berkeley.

[1] The problem of wage determination exists also in models which assume that rural employment and production is determined at the margin where utility of income and disutility of labor are equal so that marginal productivity is positive, but a gap is assumed to exist between marginal productivity in family enterprises and rural market wages. (Amartya Sen, p. 31 ff.)

I. Market Imperfections

Hanson points out that the fact, disclosed by the *EPRA*, that small farmers to a large extent find work outside their own farm does not really imply, as I contended, that "there is little reason to believe that the value of the marginal product of labor even on the smallest farms, would be smaller than current rural wages" (1969, p. 302). Preferences for self-employment may make farm families work on the farm to a point where marginal productivity is lower than current market wages. I agree in that. What makes it difficult to believe that marginal productivity generally should be lower than market wages even on small farms, is rather the fact that even small farms hire labor. We probably agree that when labor is actually hired it is hard to believe that marginal productivity should be lower than the market wage rate. Since even the smallest farms to some extent depend on hired labor (1969, p. 302, Table 3) there must be at least some time of the year when marginal productivity and market wages are in line with each other. Outside the periods when labor is hired, the preference for self-employment may, however, lead to a lower marginal product of labor than the market wage rate. The problem is then how important such preferences are. I still believe that the fact that up to 33 percent of men's work (namely on farms of $\frac{1}{2}$ to 2 feddan with more than 3 working family members) is outside their own farm, indicates that such preferences cannot be too difficult to overcome.

Another "model" (the Lewis model) mentioned by Hanson, which leads to comparison between market wage and average rather than marginal product, overlooks the fact that in an environment like rural Egypt, a decision to take temporary work outside the household, is a family decision (or, rather, a decision by the family head) and not the decision of the individual family member; for that reason it is likely that the decision is taken on the basis of the effect on *total* family income, and in that case it is, of course, the comparison between market wage and marginal product that matters.

Let it finally be pointed out that even though preference for self-employment may imply that farms that do not rely on hired labor may have a lower marginal productivity of labor than current rural market wage rate, this does not in any way upset traditional demand-supply theory (Sen, p. 33); there will still be equality between wages and the marginal product of hired labor, or, to put it another way, market wages will be equal to marginal product plus possible "marginal self-satisfaction."

II. Estimated Underemployment

With respect to estimated underemployment, and thus surplus labor, two methods should be clearly distinguished: actually observed employment may be compared with either some social norm for employment, or the "intended" supply at the going return to labor or market wages. Only the first method was applied in the case of the *EPRA*, and conceptually the second method may not even be well-defined in a market with demand-supply determined wages. (See below.)

The norm adopted by the *EPRA* was 8 hours work per day with one weekly day of rest and a few religious feast days off. It implies 303 eight-hour days per year and average annual underemployment for adult males in agriculture of about 6 percent. Hanson considers this norm too low and calculates his own underemployment percent on the basis of 306 working days (if I have understood him correctly) with a daily number of hours equal to the actual average at the seasonal peak; on this basis the average underemployment percent for the year becomes about 10 to 11. Hanson accuses the norm of the *EPRA* of being based on Western industrial practices rather than non-Western peasant agriculture. Social norms are, of course, always arbitrary. The norm chosen by the *EPRA* was largely based on what was considered to be the social norm of the present Egyptian regime, which in 1962 introduced a 42 hour week with some paid vacation in the "organized" sector and generally respects all religious feasts. From a planning point of view it is natural to consider the actual norms of the government of the country in question. The weekly day of rest (modern Western industrial practice is to have 2 such days) is, incidentally, an ancient rule, laid down in more than one major religion, thus presumably expressing the social norms of primitive non-Western peasant or tribal societies. In addition, it was taken into consideration that the average number of hours actually worked per day by hired, actually employed rural laborers turned out to be rather constant during the year June 22, 1964, to June 21, 1965 at a level of 8.1 hours. Hanson's norm seems to be based on the assumption of a given constant institutional wage rate and a horizontal supply curve (at least) to the point of actually observed peak employment. We shall see below that there is no justification for Hanson's assumption of a constant wage rate; the available evidence proves the opposite. In addition I would like to stress that even on Hanson's norm, underemployment in Egyptian agriculture is a purely seasonal phenomenon and much more limited than has usually been assumed. (The first Egyptian 5 year plan worked with the assumption that almost 25 percent of the rural labor force could be permanently removed from agriculture.)

In a further effort to bring up the average underemployment percent, Mr. Hanson reveals a peculiar, though by no means uncommon, prejudice against certain types of rural activities. I quote: "Throughout the country a large percentage of working time during [the off-season] is spent on nonagricultural and nonfarm jobs such as animal watching, marketing, and production of handicrafts . . . that have an obviously low productivity. It seems unlikely that much time could be spent in this way if, as Hansen contends, wage employment existed. If more remunerative or productive jobs were made available in the off-season, this low level work might be shifted to women or children, reducing their unemployment" (p. 495). Quite apart from the fact that "animal watching" (which is not a special Egyptian form of rodeo with which the rural population entertains itself during slack seasons, but simply animal husbandry which goes on all the year round) already to a large extent is the work of women and children, I do not understand when this kind of work has become "nonagricultural." Nor do I understand on what grounds animal husbandry, marketing, and domestic handicrafts (which in Egypt mostly means repair work and house building—rural families in Egypt usually build their mudbrick dwellings themselves) are deemed "low level work" with an "obviously low productivity," which ought to be left to women and children unless, of course, Hanson has in mind the fact that off-season marginal productivity of field work is relatively low and that any off-season activity should be expected to have low marginal productivity. Hanson does not seem to have understood that one of the important results of the *EPRA* is to demonstrate that in concentrating on field work only, earlier studies of rural employment missed a number of time consuming activities without which a traditional Egyptian farm simply cannot be run. Animals (some of them even necessary for the field work) have to be taken care of, products have to be sold and inputs purchased, and both transported, etc. And it is good to have a house to live in, even in the splendid Egyptian climate. If alternative high-productivity off-seasonal activities were created (requiring investments, presumably) farmers would, of course, shift away from both nonagricultural *and* agricultural work. But what has this to do with the problem at hand?

III. Rural Wages and Their Seasonal Behavior

Considering the great variability of daily wages, sex-age differentials between daily wages, etc., it is impossible in the collected data on daily wages to find any kind of constancy which could be interpreted as a nonmarket determined institutional wage level. Hanson believes, however, and this is his basic reinterpretation of the data, that seasonal variability of the daily wage rates is the outcome of a similar variability of the number of hours worked daily by employed laborers at a relatively constant hourly wage rate. And in this constant hourly wage rate, Hanson believes to have found the missing institutional wage rate.

This interpretation should not be ruled out a priori, of course. The *EPRA* were, however, fully aware of this possibility. Information was therefore collected not only about daily wages, but also about the number of hours to be worked at the reported daily wage. Parallel with the weekly series for daily wages, the *EPRA* thus had at its disposal a series for hours to be worked and this series turned out to be very constant with at most a very slight seasonal movement. I mentioned this in a footnote to my article (1969, p. 309) which Hanson has noticed, but misunderstood. He seems to believe that I arbitrarily assumed that number of hours for employed laborers is constant, and that he just as well can make the opposite assumption. My assumption was, however, based on firm evidence which Hanson has overlooked completely. He did not trouble himself by looking up the reference given in the footnote just mentioned. The raw weekly averages for both daily wages and daily hours for men can be found in the *EPRA Report D*, p. 39: see also graph I:4a, and figures for calculated hourly wages were published in the same volume, p. 37. The text of *Report D* discussed the problem in some detail:

> Thirdly, an important point: as an expression for the demand for labor we have used the number of hours worked on average per eight day period per member of the labor force, while as an expression for the price of labor we have preferred to use the *daily* wages, partly because the information of daily wages may be more reliable than that of hourly wages. The parallel movements of wages and hours worked per eight day period might therefore simply express the same movements for hours worked for *actually* employed people at a constant wage per hour. Now, our earlier findings actually showed a maximal seasonal spread of wages per hour of the same size as that for daily wages, and a much smaller maximal spread for hours for actually employed laborers (see Tables I:1 and I:7). It seems unlikely, therefore, that the common fluctuations in hours worked for actually employed persons [read:laborers] should explain the correlation between daily wages and average hours worked per eight day period for all members of labor force. But hours per day for actually employed laborers do show some fluctuations, and in order to make sure that the seasonality of hours worked for actually employed [laborers] does not account for the correlation between daily wages and hours worked per eight day period for all members of the labor force, we have for ***men*** calculated the average number of hours worked for actually employed [laborers], and through simple division of average daily wages and average hours actually worked the average hourly wages were obtained week by week. The upward trend in hourly wages was eliminated in the same way as for daily wages, and the "true" seasonal fluctuations of hourly wages were then obtained as the difference between the trend and the actual weekly averages. See graph I:4a. The resulting adjusted hourly wages have been depicted in graph I:5 and it turns out that the index of (adjusted) hourly wages follows the index of (adjusted) daily wages very closely. This shows that movements in the average hours actually worked per day are mainly erratic; looking at the curve for the weekly averages of hours per day in graph I:4a, one also gets the impression that there is at most a very weak seasonality in the movements of hours per day. [pp. 29–30]

To permit the reader to judge for himself, I have amended Figure 1 of my 1969 article so that it now shows not only daily wages, and average working hours (per 8 day periods) for all men in labor force, but also the average working hours (per 7 days, to make the levels easily comparable) for actually employed laborers. The difference between the two employment curves is striking. Hanson's mistake was thus to overlook that information about hours for actually employed laborers did exist, and to identify hours worked by actually employed laborers with the average of hours worked by all men in the labor force whether employed or unemployed, whether self-employed or employees. The reader will also understand now that I was fully justified in using the latter series as an index of labor demand to explain the fluctuations in daily (and incidentally, hourly) wages.

Quite apart from the fact that hourly wages thus fluctuate practically speaking in the same way as daily wages, Hanson offers us no explanation of why hourly wages should be constant and does not tell us which "institution" is implied. It would also be very queer if institutions, or, traditions were related to hourly wages, a wage form which is not and has never been used in rural Egypt. Labor is hired per day (or season, in the case of permanently employed hands), and it is significant that when the government in 1952 made an abortive attempt to institute statutory minimum wages in rural districts, it was a question of a minimum rate for *daily* wages.

If Hanson thus wants to challenge my interpretation of the *EPRA* data, he has to criticize the series for hours worked by actually employed laborers. This series is, indeed, open to some doubt. Strictly speaking it expresses the village sheik's opinion about hours to be worked at the daily wages of the week in question. He may be wrong and hours actually worked may differ from what has been agreed upon. I recall that when the *EPRA* was planned, I had expected the work time of employed laborers to fluctuate more than appears from the collected figures, not because of the fluctuations in demand for labor, but simply because the day is longer during the summer than during the winter. As in all old-fashioned agriculture the work day for hired labor in rural Egypt tends to be determined by the length of the day, and even at these southern latitudes the length of the day varies substantially during the year. Hence, my own expectation of both longer and greater fluctuation in hours of work per employed laborers.

There is no direct way of testing the reliability of the sheiks through the published material, but the following considerations may perhaps be taken as an indication that the sheiks knew what they were talking about. The statistical tables to the *EPRA* (I, Part I, Table 17, pp. 58–59) contain a frequency distribution of days by number of working hours, sex-age, and main occupation. Farm laborers are here separated out as a special group. Let us now define farm laborer days with one daily hour of work and more as "employed days." For male, employed farm laborer days, thus defined, the average number of hours work is 8.23, with a standard deviation of 1.78 hours, that is 22 percent measured upon the mean. For all males in the sample, including employed as well as unemployed days, the average number of hours is 6.33 with a standard deviation of 3.89 hours, that is 61 percent of the mean. Since our problem is to what extent the fluctuations of the average number of hours by all males (per 8 day period) can be assumed to express the fluctuations of the average number of hours for actually hired laborers, these percentage standard deviations are obviously relevant; they indicate a much larger variability in hours worked by all males than by employed laborers. However, the frequency distributions on which they are calculated are influenced by other circumstances than seasonality, for instance regional differences, farm and family size, etc. The measured standard deviations are thus larger than would follow from seasonality alone. And they are, of course, much larger than the standard deviations calculated on the figures behind the series shown in Figure 1, because on a given day in the year, all men do not work the same number of hours; the distributions of men by number of hours worked on a specific day are not available. If, however, regional differences[2] and farm and family size influence the distributions for farm laborers (which may have land, up to one-half feddan) and all men in the same way, the influence from

---

[2] I am not given access to the primary material any longer, but my recollection is that there were some differences between villages in working hours of hired labor.

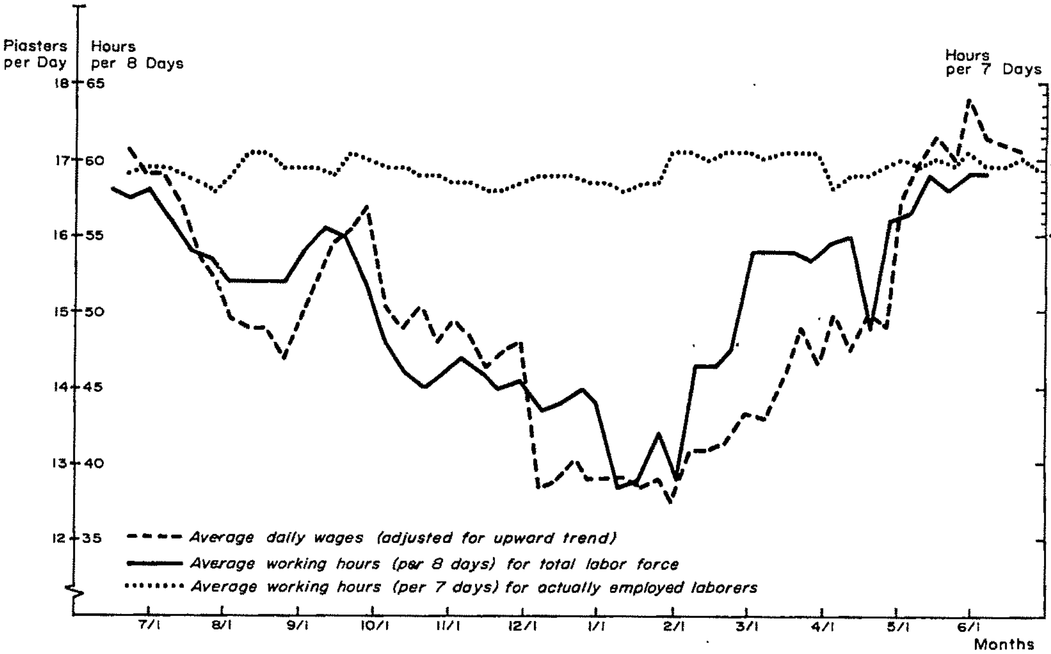


FIGURE 1

seasonality would also be in relation to the measured standard deviation. Now the average hours worked per 8 day period for all males is 50.91 and the standard deviation is 5.90 hours, that is 12 percent of the mean, which is about one-fifth of the percentage standard deviation of the frequency distribution of days by hours worked for all males, 61 percent. If the standard deviation of average hours per 8 day period for hired, actually employed laborers also were one-fifth of the standard deviation of days distributed by hours, which was 22 percent, one should expect the former standard deviation to be about 4 percent. The standard deviation of the sheik series is about 2 percent! And there is one factor which biases the measured standard deviations of days distributed by hours worked against the hypothesis of relatively constant working hours for hired laborers. As just mentioned, farm laborers may own or rent small plots of land on which they occasionally work; indeed, men in so-called landless agricultural families spent 31 and 27 percent of their work time on their own land (in small and big families, respectively). For laborers it seems likely, therefore, that days with very few hours have been spent on their own plot, and that days with very long hours have been spent partly as hired labor, partly on their own plot. If we could separate those hours which laborers have worked as hired labor, and this is what is relevant for our problem, we would probably find an even smaller percentage standard deviation for this group. No definite conclusions can be drawn on this basis, but it points at least in the direction of a much smaller variability during the year for average hours worked by actually hired labor than for all males in the sample.

Considering habits of life in Egyptian villages, I believe also now that about 8 hours effective work throughout the year may sound reasonable. Villagers rise and go to "bed" with the sun. At sunrise they get up and after a meal they are ready to walk or perhaps ride to the fields. Effective work may thus begin quite some time after sunrise. At midday there is always a long siesta and in the late afternoon, work has to stop early enough for laborers to reach their homes and get a meal before it gets dark. If we add that the midday siesta during the summer, when temperatures in the shade may increase to 120 degrees Fahrenheit, has to be considerably longer than during the winter, I think we can understand why the effective work day for laborers tends to be roughly the same all the year round. If we shall speak of nonmarket determination of conditions of work for laborers in Egyptian agriculture, the work day for laborers may be the only case in point.

Hanson has three other objections to my analysis of rural wages. The first one is related to villages with missing information for the slack season. He argues on the one hand that to leave out these villages from the average implies a bias against the demand-supply theory (presumably because thereby the measured variation in the seasonal fluctuations diminishes?) while, on the other hand, missing information discloses missing employment opportunities. The first argument is wrong, and the second does not upset the demand-supply theory, which really, as I shall show in the next section, is the only way to explain why information sometimes should be missing in the slack season.

The second objection is more valid. An analysis of wages without regard to possible regional, and Hanson might have added, seasonal differences in costs of living may go wrong in that money wage differentials may simply reflect differences in costs of living. This argument is in principle correct in so far as regional differentials are concerned. I know of no systematic study of regional differences of costs-of-living for Egypt, but there may exist significant differences between the areas covered by the *EPRA*. Considering, however, the fact that a substantial number of items entering the budget of the lower income brackets in rural areas are subject to price controls with uniform prices, I doubt very much that cost-of-living differentials can explain much of the substantial regional wage differentials. But the problem ought to be studied and I may be wrong. Differences in cost of living cannot, on the other hand, have anything to do with the strong variations in the sex-age differentials, and the seasonal fluctuations in costs of living have in fact been studied. There is quite a natural tendency for the cost of living to have a seasonal trough during the summer when vegetables and fruits are in good supply; if anything, the seasonal fluctuations of real wages are thus slightly stronger than those of money wages (see Hansen, 1966, pp. 384–85).

Hanson's last objection is related to my suggestion of using the distribution of villages by average daily wages as a test of the two wage theories. To my eye the distributions seemed to support demand-supply theory rather than institutional wage theory, but partly because of the problem with cost of living, partly confronted with his high-powered statistical test, I agree with him that no definite conclusions can be drawn by this method, although I do not quite see why he insists upon leaving out the six lowest observations.

IV. Rural Employment and Demand-Supply Theory

I see no reason thus to give up the application of demand-supply theory of wages as expressed in the regressions on page 310 of my article. It should be stressed that the theory is at variance with the application of corresponding theory to developed countries (the Phillips curve). In the application to rural wages in Egypt, I assumed that wages adjust instantaneously, which means that the labor market is always cleared and no involuntary unemployment or underemployment can exist. The point is important because it implies that if at all we shall speak about unemployment and underemployment it must be in relation to some social norm. It may be useful to show how the theory works for slack seasons in villages where no laborer finds employment, and, consequently, it seems hard to insist that the market is cleared, and to deny that there is unemployment.

The following simple model of rural employment and wage formation is an adaptation of Chihiro Nakajima's model for family farms (see in particular, pp. 176 and 179–81). Nakajima considers only the case when the wage rate is given exogenously; our problem is the determination of both employment and wages. For the sake of simplification I disregard work of farm family members outside the farms (small farmers working temporarily for big farmers, or for government) but the model can easily be extended to cover that case too.

Assume thus that we have farm families working on their own land, and landless laborers working for farm families. In Figure 2A, I have depicted the aggregate short run supply curve of farm families S^f, and landless laborers, S^l, respectively. Since the supply of farm family labor is used only for input on their own farm, I shall assume that the supply curve for farm family labor starts out from a lower level than that of landless laborers; this assumption is in line with the idea that self-employment is considered more respectable and/or enjoyable than work for others. Since we are considering the very short run, supply becomes zero at a positive wage rate. The total supply curve for labor is thus S. In Figure 2B we have the "open" or "net" labor market; labor supply in the open market is S^l. We make no distinction between hours and days, but this complication could easily be taken care of (see Sen).

Let then MP in Figure 2A depict the value of marginal productivity (at given farm product prices), falling with increased labor input.[3] The marginal productivity curve for a day or week at the peak season is MP_1. The corresponding net demand curve in the open market is D_1, which is the excess demand of farms, $MP_1 - S^f$ (horizontally). Utility maximization for the farm families requires that production is extended to the point where the market wage rate is equal to the value of marginal productivity, and that farm family labor is extended to the point where the farm family labor supply price is equal to the wage rate. At MP_1, total labor input becomes L_1, family labor input L_1^f and input of hired labor L_1^l. The wage rate is

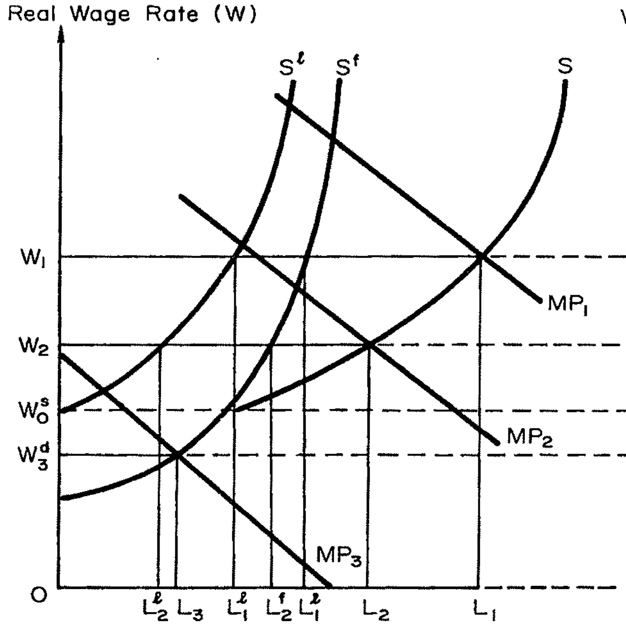


FIGURE 2A

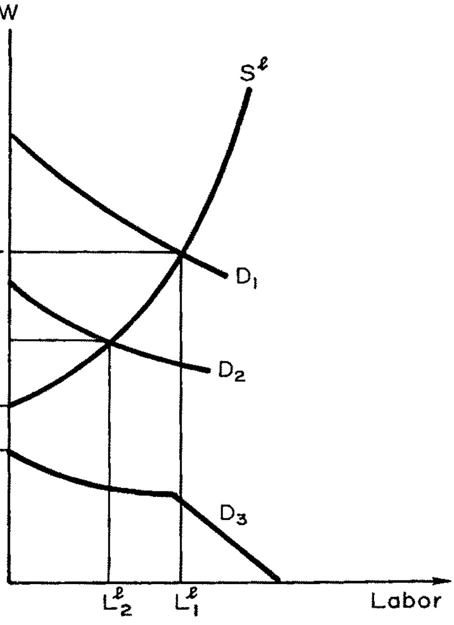


FIGURE 2B

---

[3] Let $Q_{j\tau}$ be the output of product j maturing for sale at time τ, and I_{it} be the input of factor i at time t, $t<\tau$. There are m products and n factors. The value of the marginal productivity of factor i at time t is then

$$MP_{it} = \sum_{\tau=t}^{\infty}\sum_{j=1}^{n} \frac{p_{j\tau}}{(1+r)^{\tau-t}}\cdot\frac{\partial Q_{j\tau}}{\partial I_{it}}$$

where $p_{j\tau}$ is the price of output j expected for time τ, r the interest rate per period (say, day), and the derivatives are calculated on the transformation function $F(Q_{j\tau}; I_{it})=0$. The optimum condition is $W_{it}=MP_{it}$. Clearly MP will shift up and down during the year with the biological and climatic rhythms of agricultural production.

W_1. At a less pronounced peak, the marginal productivity curve may be MP_2 and the net demand in the open market, D_2. Both family and hired employment will be lower, and so will the market wage rate which is now W_2. We note that no involuntary unemployment or underemployment have arisen; both farm families and laborers work at their optimum. Consider then a slack season with MP_3. Here farm family employment has fallen to L_3 and no hired labor is employed. Wages are *indeterminate.* The supply price for laborers at zero employment is W_0^s, while the demand price at L_3 is W_3^d. We have thus a situation where an equilibrium in the open labor market does not exist; no laborers are hired and there exists no market wage rate.[4] Marginal productivity is positive, but it could fall to zero without changing anything in the open market situation. If the MP curve passed through the origin (in which case not even family members would work), or the S^f-curve in the beginning coincides with the x-axis (in which case family members may do some work even at a very low marginal productivity curve), the supply price of laborers would still be W_0^s at a demand price equal to zero and the market wage rate would still be indeterminate. Even in this position we cannot determine the amount of involuntary unemployment, because there is no definite market wage rate at which supply could be measured.

The pattern described here corresponds closely to what actually happened in the wage survey. As the season became more slack, both wages and employment fell; and in some villages, before wages fell to zero, there were suddenly no wages to report; there was no active labor market, no hiring of labor. These are the villages with information about wages missing for the slack season. If there had existed an institutionally, or traditionally given wage rate, there is no reason why it should be impossible for the sheik to tell us what the wage rate is when there is no hiring of labor.

V. Concluding Remarks

Since there is full agreement about the existence of some seasonal underemployment and unemployment, in particular in Upper Egypt, Hanson's concluding recommendations of counterseasonal employment policies can give rise to no disagreement either. It has always been recognized that the perennial cropping which now is being substituted for basin irrigation in parts of Upper Egypt will help to wipe out seasonal underemployment there. The main problem is how the electricity from the Asswan High Dam will be used. It seems now that priority will be given to village electrification which is probably a condition for creating counterseasonal employment in the villages and for mechanization of agriculture (repair shops). On the other hand, it should not be overlooked that to the same extent as seasonal underemployment of labor is diminished in this way, seasonal underemployment of capital may be created. It is not obvious what is best from a social point of view.

REFERENCES

B. Hansen, "Employment and Wages in Rural Egypt," *Amer. Econ. Rev., 59,* June 1969, 298–313.

———, "Marginal Productivity Wage Theory and Subsistence Wage Theory in Egyptian Agriculture," *J. Develop. Stud., 2,* July 1966, 367–407.

C. Nakajima, "Subsistence and Commercial Family Farms: Some Theoretical Models of Subjective Equilibrium," in C. R. Wharton, Jr., ed., *Subsistence Agriculture and Economic Development,* Chicago 1969, pp. 165–85.

A. K. Sen, "Peasants and Dualism with or without Surplus Labor," *J. Polit. Econ.,* Oct. 1966, *74,* 425–50.

Institute of National Planning, *Employment Problems in Rural Areas,* U.A.R.

(I) *Statistical Tables, The Labour Record Sample Survey,* Part I, Cairo, Dec. 1965; Part II, Cairo, Sept. 1966.

(II) *Report D on Wages, Income, and Consumption in Rural Areas,* prepared by B. Hansen, A. Sedki, and Y. Moustafa, Cairo, Dec. 1965.

---

[4] What prevents a solution from existing is a discontinuity in the excess demand function. The proofs for the existence of a competitive equilibrium generally assume continuous excess demand functions.

ERRATA

The Determinants of U.S. Direct Investment in the E.E.C.

By ANTHONY E. SCAPERLANDA AND LAURENCE J. MAUER

A number of errors occurred in Tables 1 and 2 of this article which appeared in the September 1969 *Review*, pages 558–68. The errors arose from the use of a regression program employing the single precision computational technique and in one instance, as the result of a scale change.

The corrected tables follow:

TABLE 1—REGRESSION EQUATIONS FOR *U.S.* DIRECT FOREIGN INVESTMENTS IN THE *E.E.C.*: ANNUAL DATA, 1952–1966[a]

	Regression Equation	$\bar{R}^2$	S_e	DW
(1.1)	$I=-426.707+5.520Y-58.860M+349.407\Delta M-8.099\Delta Y$ (5.171)** (.125) (.821) (.765)	.952**	79.697	1.665**
(1.2)	$I=-362.665+4.817Y-125.006M+337.617\Delta M-386.487G_1$ (9.925)** (.248) (.774) (.239)	.950**	81.763	1.572
(1.3)	$I=-364.682+4.764Y-188.711M+346.609M-.036G_2$ (8.443)** (.349) (.702) (.041)	.950**	81.989	1.570
(1.4)	$I=-463.272+4.899Y-.763M$ (12.129)** (.002)	.955**	77.045	1.572*
(1.5)	$I=-461.624+4.911Y+252.639\Delta M$ (17.783)** (.725)	.957**	75.410	1.498**
(1.6)	$I=-459.966+5.430Y-6.605\Delta Y$ (6.872)** (.716)	.957**	75.451	1.720*
(1.7)	$I=-436.758+4.899Y-293.815G_1$ (17.424)** (.221)	.956**	76.889	1.599*
(1.8)	$I=-471.167+4.928Y+.217G_2$ (16.762)** (.321)	.956**	76.716	1.611*
(1.9)	$I=-372.011+4.778Y-176.763M+337.201\Delta M$ (10.941)** (.406) (.809)	.954**	78.180	1.572**
(1.10)	$I=-518.935+5.616Y+118.506M-7.773\Delta Y$ (5.373)** (.288) (.746)	.954**	78.510	1.670**
(1.11)	$I=-454.159+4.938Y+50.536M-381.464G_1$ (10.953)** (.114) (.240)	.952**	80.261	1.577**
(1.12)	$I=-496.358+4.970Y+47.915M+.248G_2$ (10.548)** (.116) (.329)	.952**	80.079	1.592**
(1.13)	$I=-456.274+5.600Y+324.932\Delta M+8.533\Delta Y$ (6.842) (.901) (.894)	.957**	76.048	1.645**
(1.14)	$I=-410.177+4.912Y+289.040\Delta M-558.764G_1$ (17.150)** (.776) (.400)	.954**	78.197	1.528**
(1.15)	$I=-466.140+4.927Y+239.530\Delta M+.127G_2$ (16.347)** (.646) (.180)	.954**	78.647	1.526**
(1.16)	$I=-463.656+4.900Y$ (18.104)**	.959**	74.022	1.571*

[a] The numbers in parentheses are *t*-statistics; * and ** indicate significantly different from zero at 5 percent and one percent levels, respectively.

The symbol * and ** on the Durbin-Watson statistics indicates that the null hypothesis of residual independence cannot be rejected at the 5 percent or one percent level of significance. The tests of the *DW* statistics are based on the H. Theil–A. Nager testing procedure.

TABLE 2—REGRESSION EQUATIONS FOR *U.S.* DIRECT FOREIGN INVESTMENTS IN THE *E.E.C.*: ANNUAL DATA[a]

Regression Equation	$\bar{R}^2$	S_e	*DW*
1952–58:			
(2.1) $I=-245.359+3.294Y$ (5.109)**	.807**	36.024	1.845
(2.2) $I=-339.707+2.736Y+399.066M$ (5.520)** (2.553)	.908**	24.838	2.438
(2.3) $I=-243.567+3.303Y+189.690\Delta M$ (5.219)** (1.092)	.814**	35.350	2.125
(2.4) $I=-245.903+3.060Y+2.904\Delta Y$ (2.954)** (.312)	.765**	39.796	1.954
(2.5) $I=-260.521+3.299Y+161.522G_1$ (4.594)** (.192)	.761*	40.092	1.905
(2.6) $I=-232.603+3.217Y-0.153G_2$ (4.417)** (.427)	.769**	39.389	1.919
1959–66:			
(2.7) $I=-714.528+5.871Y$ (11.064)**	.945**	75.069	2.629
(2.8) $I=-1069.750+6.470Y+678.706M$ (4.003)** (.396)	.937**	80.971	2.493
(2.9) $I=-722.036+5.976Y+1018.147\Delta M$ (11.026)** (.990)	.945**	75.196	2.550
(2.10) $I=-699.561+6.118Y+3.696\Delta Y$ (6.079)** (.300)	.936**	81.503	2.747
(2.11) $I=-690.903+5.859Y-225.577G_1$ (9.820)** (.087)	.935**	82.172	2.665
(2.12) $I=-697.549+5.835Y-5.292G_2$ (8.750)** (.107)	.935**	82.139	2.642

[a] See note to Table 1 for explanation of * and **. Numbers in parentheses are *t*-statistics. None of the *DW* statistics were tested owing to the small number of degrees of freedom.

NOTES

Nominations for AEA Offices

The Electoral College on March 5 chose Kenneth J. Arrow as nominee for president-elect of the American Economic Association in the balloting to be held in the autumn of 1971. Other nominees (chosen by the nominating committee) are: for vice president (two to be elected), Ronald H. Coase, John G. Gurley, Hendrik S. Houthakker, and Arthur M. Okun; for member of the excutive committee (two to be elected), Gerard Debreu, Guy H. Orcutt, Joseph A. Pechman, and Melvin W. Reder.

Under a change in the bylaws reported in the *Papers and Proceedings* of this *Review*, May, 1971, additional candidates may be nominated by petition, delivered to the secretary by August 1, including signatures and addresses of not less than 6 percent of the membership of the Association for the office of president-elect and not less than 4 percent for each of the other offices. For the purpose of circulating petitions, address labels will be made available by the secretary at cost.

Announcement of Junior Dues Increase

At the annual meeting of the American Economic Association on December 29, 1970, those present voted to double the dues of all except junior members. A subsequent vote provided that the Executive Committee could raise the dues of junior members to the extent necessary to retain second class postal privileges.

The AEA counsel, Seymour J. Rubin, has ascertained that the dues of junior members need to be doubled for this purpose. The Executive Committee of the AEA voted, March 5, 1971, to increase the dues for junior members to ten dollars. An abstract of Mr. Rubin's opinion is given below:

> Part 132.228 of the "Post Office Regulations for Mail Classifications and Rates: Second Class" provides, essentially, that a publisher can charge a rate not less than 50 percent of its full subscription rate to a "premium group" of subscribers and still qualify for second class privileges, so long as the reduced rate charged is not less than the cost to the publisher of the subscription rate involved. Thus if your full subscription rate is $20, or some stated portion of the $20 membership fee you charge to regular members, you are allowed under the regulations to charge no less than half that rate ($10) to a premium group such as junior members, assuming the reduced rate is not less than your cost.

The American Academy of Political and Social Science recently held a conference, "Harmonizing Technological Developments and Social Policy in America." A monograph consisting of five principal papers, critiques on each, and proceedings of the conference may be obtained free by writing to the Academy, 3937 Chestnut Street, Philadelphia, Pennsylvania 19104.

The Committee on International Exchange of Persons announces availability of foreign economists for appointments in U.S. universities and colleges under provisions of the Fulbright-Hays Act for the academic year 1971–72.

Japan: Hiromi Ishigaki, Kaichi Shimura; Yugoslavia: Miroslav Petrovic, Branislav Soskic, Dragoljub Stojilkovic, Boris Tihi; Italy: Alessio Lokar, Alfredo Medio, Rino Ricci; Turkey: Turgent Var.

Inquiries should be addressed to: Miss Grace E. L. Haskins, Program Officer, Conference Board of Associated Research Councils, 2101 Constitution Avenue, Washington, D.C. 20418 or via telephone—Area Code 202, 961-1948.

The National Science Foundation will sponsor a summer institute for junior college instructors of economics, August 2–27, 1971. The topics will include contemporary economic analysis and the economics of education. Director: H. Robert Heller, economics department, University of California at Los Angeles, Los Angeles, California 90024.

The National Institute of Social and Behavioral Science will hold sessions for contributed papers at the 138th annual meeting of the American Association for the Advancement of Science, December 26–31, 1971, in Philadelphia. Economists who would like to give a paper at these sessions are invited to forward titles and abstracts of 300 words by August 25 to Donald P. Bay, National Institute of Social and Behavioral Science, 863 Benjamin Franklin Station, Washington, D.C. 20044.

Suggested topics for papers might include in whole or in part the economics of selected wage-price controls; a national income policy; the Federal full-employment budget in theory and practice; percentage state income tax deductions from federal income taxes as an optimum form of revenue-sharing; environmental and welfare economics; some aspects of scope, degree, and intensity in the redistribution of income in the United States; reciprocal trade agreements questions and foreign dumping as current problems in the U.S. balance of payments; employment and the status of free trade theory in a high inflation economy; automation, computers, and management science as instruments of social and economic change in the U.S.S.R.; foreign economic competition in postwar Southeast Asia and problems in the protection of economic nationalism; general international investment, trade, and exchange questions; economic development theory; and subjects in interdisciplinary areas.

The proceedings of a research conference, held Sept. 11–12, 1970 by the Inter-University Committee on Urban Economics, are now available. The papers cover urban land and housing markets, economic consequences of air pollution, poverty, public goods theory, and related policies. Orders enclosing $3.00 payable to the University of Chicago should be addressed to Professor G. S. Tolley, department of economics, 1126 E. 59th Street, Chicago, Illinois 60637.

A two-day meeting will be held at Oklahoma State University, Stillwater, on the relationships between computer science and statistics. The keynote speaker will be H. O. Hartley of Texas A&M University. Workshops will be conducted in five areas: TIME SERIES AND STOCHASTIC PROCESSES: Emanuel Parzen, State University of New York, Buffalo, chairman: DECISION SCIENCES: Dennis Grawoig, Georgia State University, Atlanta, chairman: COMPUMETRICS: Robert Gordon, University of California, Irvine, chairman, COMPUTER SCIENCE AND STATISTICS IN HIGHER EDUCATION: J. L. Folks, Oklahoma State University and Ron Mohler, University of Oklahoma, co-chairmen; COMPUTER SCIENCE AND STATISTICS IN THE EXTRACTIVE INDUSTRIES.

Persons having papers in any of these five areas (up to about eight pages) are invited to submit them to the conference chairman, Dr. Mitchell O. Locks, Oklahoma State University, Stillwater, Oklahoma 74074.

The American Academy of Political and Social Science, 3937 Chestnut Street, Philadelphia, Pennsylvania 19104, announces that it will send free upon request, copies of a study "Design for International Relations Research: Scope, Theory, Methods, and Relevance."

Deaths

A. B. Andarawewa, economics branch, Canada Department of Agriculture, Ottawa, fall 1970.

Frederick A. Bradford, professor emeritus, Lehigh University, Jan. 22, 1971.

Elwyn Cady, Kansas City, Missouri, Dec. 15, 1970.

Herbert B. Dorau, professor emeritus of economics, School of Commerce, New York University, Jan. 17, 1971.

Eberhard M. Fels, University of Munich, Germany, spring 1970.

Paul Gekker, Board of Governors of the Federal Reserve System, Jan. 16, 1971.

Everett B. Hurt, assistant professor, Scarborough College, University of Toronto, Feb. 14, 1971.

Jerzy F. Karcz, professor of economics, chairman, Russian area studies, University of California at Santa Barbara; general editor, *Eastern European Economics*; Santa Barbara, Dec. 10, 1970.

Doris G. Phillips, professor of economics, California State College, Fullerton, Jan. 14, 1971.

Robert Rockafellow, professor of economics, University of Rhode Island, Dec. 16, 1970.

Erich Schneider, Germany, Dec. 5, 1970.

Gilbert Stonesifer, professor and chairman, department of economics and business administration, Mount Union College, Oct. 1970.

Clair Wilcox, professor of political economy emeritus, Swarthmore College, Dec. 31, 1970.

Retirements

Roland B. Eutsler, professor of economics, University of Florida, Oct. 1, 1970

Morris E. Garnsey, department of economics, University of Colorado, fall 1971.

Robert W. Harbeson, professor of economics, University of Illinois, Urbana, Aug. 1971.

Ben W. Lewis, International Division, The Ford Foundation, Aug. 31, 1971.

Ralph Sherman, Ohio State University, Feb. 1, 1971.

Carl S. Shoup, professor of political economy, Columbia University, June 1971.

Elbridge Sibley, Social Science Research Council, Dec. 31, 1970.

N. Arnold Tolles, Cornell University: professor of economics, New York State University at Geneseo.

Leland S. Van Scoyoc, professor of economics, Bowling Green State University, Dec. 1970.

Paul Webbink, Social Science Research Council, Dec. 31, 1970.

Visiting Foreign Scholars

H. G. Brennan, Australia National University: visiting foreign scholar, department of economics, Dalhousie University, 1971–72.

Bienvenido Delantar, Trinity College of Quezon City: visiting assistant professor in economics, Trinity College, fall 1970–Jan. 1972.

Kenneth A. Leslie, University of the West Indies, Kingston, Jamaica: Food Research Institute, Stanford University, 1970–71.

Broadus Mitchell, Hofstra University: visiting professor of economics, University of Rhode Island, spring 1971.

Luigi Pasinetti, Cambridge University: visiting professor, Columbia University, fall 1971.

Joan Robinson, Cambridge University: visiting professor of economics, University of Waterloo, fall term 1971.

Marcel M. Tardos, Institute for Economics and Market Research, Budapest, Hungary: visiting lecturer, department of economics, University of Michigan, winter 1971.

Alfred Zauberman, London School of Economics and Political Science: visiting lecturer in economics, University of California, Santa Barbara, winter, spring 1971.

Promotions

M. William Belovicz: associate professor of economics, University of Massachusetts, Feb. 1, 1971.

Nicholas G. Bohatiuk: professor of economics, Le Moyne College.

Lawrence A. Boland: associate professor, department of economics and commerce, Simon Fraser University.

E. Gerald Corrigan: chief, domestic research division, Federal Reserve Bank of New York.

Richard V. Cotter: professor in managerial sciences, University of Nevada, July 1, 1971.

Ralph C. d'Arge: associate professor of economics, University of California, Riverside.

Michael Edelstein: assistant professor of economics, Columbia University.

Samuel M. Ehrenhalt: deputy regional director, Bureau of Labor Statistics.

Stanley L. Engerman: professor of economics, University of Rochester, Sept. 1, 1971.

James Esmay: associate professor, San Fernando Valley State College.

Frederick Finch: associate professor of economics, University of Massachusetts, Feb. 1, 1971.

Raymond S. Franklin: associate professor of economics, Queens College of the City University of New York.

Martin Geisel: associate professor of economics, Carnegie-Mellon University, Sept. 1, 1971.

Louis Geller: associate professor, Queens College of the City University of New York.

Harry I. Greenfield: professor of economics, Queens College of the City University of New York.

Everett E. Hagen: director for center of international studies, Massachusetts Institute of Technology.

William Haller, Jr.: professor of economics, University of Rhode Island, July 1, 1971.

Harry J. Halley: deputy controller and program director for management systems, Board of Governors of the Federal Reserve System, Jan. 1971.

John J. Hooker: professor, department of economics and finance, University of Texas at El Paso.

Peter E. Kennedy: associate professor, department of economics and commerce, Simon Fraser University.

Leonard Lapidus: assistant vice president, Federal Reserve Bank of New York

Larry J. Larsen: professor of economics, University of Nevada, July 1, 1971.

Lester B. Lave: professor of economics, Carnegie-Mellon University, Sept. 1, 1971.

L. C. Ledebur: associate professor of economics, Denison University.

Will Lyons: professor of economics, Franklin and Marshall College.

Donald Mathieson: assistant professor of economics, Columbia University.

Kent Monroe: associate professor of economics, University of Massachusetts, Feb. 1, 1971.

Woo H. Nam: associate professor of economics, San Diego State College.

Peter Passell: assistant professor of economics, Columbia University.

Malvika Patel: associate professor, department of economics, San Fernando State College.

John D. Patrick: chief, balance of payments division, Federal Reserve Bank of New York.

James L. Pierce: associate adviser, division of research and statistics, Board of Governors of the Federal Reserve System, Jan. 1971.

A. Marshall Puckett; adviser, research and statistics function, Federal Reserve Bank of New York.

Fredric Q. Raines: associate professor of economics, Washington University, St. Louis, July 1, 1970.

Robert Rivers: professor of economics, University of Massachusetts, Feb. 1, 1971.

Dick L. Rottman: professor of managerial sciences, University of Nevada, July 1, 1971.

Gregory C. Schmid: chief, foreign research division, Federal Reserve Bank of New York.

James Starkey: assistant professor of economics, University of Rhode Island, July 1, 1971.

Rudolf Thunberg: manager, domestic research department, Federal Reserve Bank of New York.

Sheila Tschinkel: special assistant, securities department, Federal Reserve Bank of New York.

William E. Whitesell: associate professor of economics, Franklin and Marshall College.

Parker Worthing: associate professor of economics, University of Massachusetts, Feb. 1, 1971.

Administrative Appointments

Alec P. Alexander: academic assistant to chancellor, University of California, Santa Barbara, Jan. 1971.

Wallace N. Atherton: chairman, department of economics, California State College at Long Beach, Feb. 1, 1971.

M. E. Bond: director, bureau of business and economic research, Arizona State University, July 1971.

Flournoy A. Coles, Jr.: director, center for black economic development, Fisk University.

Josephine M. Corrigan: assistant to the dean, St. John's University, Sept. 1970.

Richard S. Cowan: chairman, department of economics and business administration, Waynesbury College.

Simeon J. Crowther: associate chairman, department of economics, California State College at Long Beach, Feb. 1, 1971.

Ward S. Curran: director of institutional development, Trinity College, 1971–73.

Dennis J. Dugan: chairman, department of economics, University of Notre Dame, Jan. 1, 1971.

Melvin A. Eggers: vice chancellor for academic affairs and provost, Syracuse University.

Joseph A. Giacalone: assistant dean, St. John's University, Sept. 1970.

Harold M. Hochman: director of studies in urban public finance, Urban Institute.

Kenneth C. Kehrer: acting chairman, department of economics and business administration, Fisk University.

William H. Leahy: director of undergraduate studies, University of Notre Dame, Jan. 1, 1971.

Gordon W. McKinley: senior vice president, economics and financial planning, McGraw-Hill, Inc.

Hugh S. Norton: chairman, department of economics, University of South Carolina, fall 1970.

James E. Price: chairman, economics department, Syracuse University.

G. Ray Prigge: chairman, department of economics and business administration, Mount Union College.

Clarence A. Reed: manager of developmental research, marketing research department, The Coca-Cola Company.

M. Selim: director of center for economic education, College of St. Thomas.

Roger M. Skurski: director of graduate studies, University of Notre Dame, Jan. 1, 1971.

T. Arthur Smith, U. S. Army director of cost analysis: chairman, Defense Economic Analysis Council.

Richard J. Solie: head, department of economics, University of Alaska, Sept. 1970.

Peter O. Steiner: chairman, department of ceonomics, University of Michigan.

Vito Tanzi: chairman, department of economics, The American University.

J. Vanderkamp: chairman, department of economics, College of Social Science, University of Guelph, July 1, 1971.

New Appointments

I. Robert Andrews, Bowling Green State University: associate professor, department of economics and commerce, Simon Fraser University.

Harold Barnett, Federal Reserve Bank of Boston: instructor, department of economics, University of Rhode Island, fall 1970.

Randall Bartlett: assistant professor, department of economics, Williams College.

William J. Baumol: professor of economics, New York University.

H. Prescott Beighley: economic research unit, division of research, F.D.I.C., Washington

Sanford V. Berg: assistant professor of economics, University of Florida, 1971–72.

Rudolf M. Binnewies, University of California at Berkeley: instructor, department of economics, Vassar College.

Harold A. Black: assistant professor of economics, University of Florida, 1971–72.

Richard Boisvert: assistant professor, department of economics, Cornell University.

Myles G. Boylan, Jr.: instructor of economics, Case Western Reserve University.

William A. Brock, University of Rochester: visiting associate professor of economics, University of Chicago, 1971–72.

Karl Brunner: professor of economics, joint appointment, department of economics and Graduate School of Management, University of Rochester.

John Burge: visting assistant professor, department of economics, Mercer University, Sept. 1970.

Charles Cathcart: assistant professor of economics, Pennsylvania State University, Sept. 1971.

Don M. Chaffee: assistant professor of economics, University of South Carolina, fall 1970.

Divakar J. F. Chandran: assistant professor of economics, University of Alaska, Sept. 1970.

L. Duane Chapman, University of Tennessee: assistant professor, department of economics, Cornell University.

Henry E. Cole, Johns Hopkins University: assistant professor of economics, Tulane University.

Flournoy A. Coles, Jr., Fisk University: professor of economics, Graduate School of Management, Vanderbilt University.

Steven G. Darr: instructor, department of economics and business administration, Mount Union College.

Martin H. David, University of Wisconsin: professor of economics, University of Michigan.

J. Ronnie Davis: associate professor of economics, University of Florida, 1971–72.

Robert M. Deaver: lecturer in economics, Arizona State University, spring 1971.

Rudiger Dornbusch: assistant professor of economics, University of Rochester, Sept. 1971.

Ernest G. Ellingson: assistant professor of economics, Ball State University.

Dennis Ellis, Wayne State University: assistant professor, department of economics, Southern Illinois University.

Asim Erdilek: instructor of economics, Case Western Reserve University.

Edward C. Ettin: assistant adviser, division of research and statistics, Board of Governors of the Federal Reserve System, Jan. 1971.

Robert Falconer: economist, financial statistics division, Federal Reserve Bank of New York.

Terry A. Ferrar: assistant professor of economics, Pennsylvania State University, Sept. 1971.

Ray W. Fowler: professor of buisness, department of economics, Walla Walla College, July 1970.

Mark W. Frankena: lecturer, department of economics, University of Western Ontario.

Richard J. Gelson: economist, financial statistics division, Federal Reserve Bank of New York.

Gary G. Gilbert: economic research unit, division of research, F.D.I.C., Washington.

Peter Gregory, University of Minnesota: professor of economics, University of New Mexico.

Klaus D. Grimm, educational testing service, Princeton: research associate, department of public social services, county of Marin, California.

Herbert G. Grubel, University of Pennsylvania: professor, department of economics and commerce, Simon Fraser University.

Peter E. Gunther: assistant professor of economics, Mount Allison University.

Demos Hadjiyanis: associate professor of economics, College of St. Thomas.

Peter Heller, Harvard University: assistant professor, department of economics, University of Michigan.

John Hill: economist, domestic research division, Federal Reserve Bank of New York.

A. G. Holtmann, University of Wisconsin; professor of economics, State University of New York at Binghamton.

Yutaka Horiba: assistant professor of economics, Tulane University.

George C. Hoyt, University of Iowa: professor, department of economics and commerce, Simon Fraser University.

David B. Humphrey: associate professor of economics, Tulane University.

Thomas O. Jones, Jr.: professor of economics and management, U.S. Coast Guard Academy, fall 1970.

Harry Kelejian: professor of economics, New York University.

Clifford R. Kern, Harvard University: assistant professor of economics, State University of New York at Binghamton, Sept. 1971.

Karen I. Kidder: economist, banking studies department, Federal Reserve Bank of New York.

Young Kim: assistant professor, department of economics, Northern Illinois University.

Tetteh A. Kofi: visiting professor, Food Research Institute, Stanford University, 1970–71.

L. Emil Kreider: assistant professor of economics, Beloit College, fall 1970.

Peter J. Kuch: lecturer, department of economics, University of Western Ontario.

Robert Kuller, University of Kansas: assistant professor, department of economics, University of Wyoming.

Leonard Kunin: associate professor of economics, San Jose State College, fall 1971.

Roslyn Kunin: visiting assistant professor, department of economics and commerce, Simon Fraser University.

Abba P. Lerner, University of California, Berkeley: professor of economics, Queens College of the City University of New York, 1971–72.

Roger D. Little, University of Houston: assistant professor, economics department, U.S. Naval Academy, Aug. 1970.

Peter C. Lin: assistant professor of economics, University of Alaska, Sept. 1970.

Ronald G. Lorentson: acting assistant professor of agricultural economics and acting assistant economist, experiment station, University of California, Berkely, fall 1970.

Jerolyn R. Lyle, U.S. Equal Employment Opportunity Commission: lecturer in economic theory, Smith College, spring 1971.

Michael D. McCarthy: associate professor, department of economics, Case Western Reserve University, Sept. 1971.

Thomas O. McCoy: assistant professor, department of economics, Williams College.

Donald G. McFetridge: lecturer, department of economics, University of Western Ontario.

Peter F. M. McLoughlin, University of California, Santa Clara: visiting professor of economics, department of political economy, University of Toronto, Sept. 1970.

James M. Mangum, Oklahoma State University: associate professor of economics, Louisiana Tech University, Sept. 1970.

Harry Mapp: assistant professor, department of economics, Cornell University.

Laurence H. Meyer: assistant professor of economics, Washington University, July 1, 1969.

J. Robert Moore, Millersville State College: assistant professor of business administration, University of Wisconsin, Feb. 1, 1971.

Yasuo Murata: professor, department of economics, Dalhousie University, July 1, 1971–June 1973.

Michael Mussa: assistant professor of economics, University of Rochester, Sept. 1971.

Soren T. Nielsen: instructor, department of economics and commerce, Simon Fraser University.

Richard B. Norgaard: acting assistant professor of agricultural economics and acting assistant economist, experiment station, University of California, Berkeley fall 1970.

Ronald L. Oaxaca: visiting lecturer, department of economics, University of Western Ontario.

Thomas R. O'Connor, State University of New York at Buffalo: instructor, department of economics, Wabash College.

E. Odgers Olsen, University of North Carolina: assistant professor of economics, State University of New York at Binghamton, Sept. 1971.

John P. Palmer: lecturer, department of economics, University of Western Ontario.

Alan Pankratz: assistant professor of economics, DePauw University, 1970–71.

Clifford H. Patrick: economic research unit, division of research, F.D.I.C., Washington.

Prasanta K. Pattanaik: associate professor, department of economics, University of Western Ontario.

Nicholas Perna: economist, domestic research division, Federal Reserve Bank of New York.

Bruce Petersen, National Bureau of Economic Research: instructor, department of economics, University of Wyoming.

Manferd O. Peterson: economic research unit, division of research, F.D.I.C., Washington.

John T. Pickerill: assistant professor of economics, Ball State University.

James P. Quirk: professor of economics, division of the humanities and social sciences, California Institute of Technology, fall 1971.

Robert Rice: assistant professor of economics, University of Hawaii.

Sergio Roca: instructor, department of economics, Adelphi University.

Richard K. Rudel: assistant professor, economics department, South Dakota State University, Feb. 1, 1971.

William R. Russell, University of Kentucky: professor of economics, Southern Methodist University.

Stephen R. Sacks: assistant professor of economics, University of Connecticut.

David T. Scheffman: visiting lecturer, department of economics, University of Western Ontario.

Peter J. Schmidt: assistant professor, department of economics, Wayne State University, Jan. 1, 1971.

Robert N. Schoeplein, University of Connecticut: associate professor, department of economics and institute of government and public affairs, University of Illinois, Urbana.

William L. Scott, Jr.: instructor, economics and business, Mercer University, Sept. 1970.

David E. Seckler, University of California, Berkeley: professor, department of economics, Colorado State University, Mar. 25, 1971.

Lawrence C. Smith, Mississippi College: professor of economics, Louisiana Tech University, June 1970.

Robert S. Smith: assistant professor of economics, University of Connecticut.

Norman H. Starler, South Dakota State University: assistant professor of economics, State University of New York College at Fredonia.

William Stull; assistant professor of economics, Swarthmore College.

John Swirles, University of British Columbia: visiting associate professor, department of economics and commerce, Simon Fraser University.

Leslie Szeplaki: assistant professor, department of economics, Northern Illinois University.

J. Ernest Tanner, University of Western Ontario: associate professor of economics, Tulane University.

Reed Taylor: department of agricultural economics and rural sociology, Ohio State University, Jan. 1, 1971.

Thomas H. Tietenberg: assistant professor, department of economics, Williams College.

Ronald L. Tinnermeier, North Carolina State University: associate professor, department of economics, Colorado State University 1971–72.

Tore Tjersland: professor, department of economics, Colorado State University, 1971–72.

Lynn Turgeon: visiting professor, department of economics, San Jose State College, fall 1971.

Susan Wachter: lecturer in economics, department of economics, Bryn Mawr College, 1971–72.

William Walsh: assistant professor of economics, College of St. Thomas.

David Weinberg, University of California at Berkeley: instructor, department of economics, Vassar College.

Roger S. White: assistant professor of economics, University of Connecticut.

Ronald Wilder, Vanderbilt University: assistant professor, department of economics, University of South Carolina, fall 1970.

Robert S. Woodward: lecturer, department of economics, University of Western Ontario.

Gregory J. L. Yi: assistant professor of economics, West Virginia University.

Leaves for Special Appointments

Martin Bronfenbrenner, Carnegie-Mellon University: visiting professor of economics, Aoyama-Cakuin University, Tokyo, fall 1970.

William E. Gibson, University of California at Los Angeles: Brookings Economic Policy Fellowship, economic research unit, division of research, F.D.I.C., Washington.

Philip D. Grub, School of Government and Business Administration, George Washington University: visiting professor of international marketing, economics faculty, University of Helsinki, spring semester 1971.

Woo Sik Ki, West Virgina University: research department, Korea Exchange Bank.

Arlyn J. Larson, Arizona State University: legislative research assistant, Arizona Legislature, spring 1971-summer 1972.

Huntley G. Manhertz, University of Pittsburgh: Institute of Social and Economic Research, University of the West Indies.

David C. Warner, Wayne State University: senior economist, Health and Hospital Corporation of New York City.

Resignations

Leslie Aspin, Marquette University: Representative, 1st Congressional District, Wisconsin.

Robert W. Baguley, University of Western Ontario, June 1971.

Thomas J. Beirne, University of Nevada, May 1971.

John A. Dominick, Louisiana Tech University: University of Arkansas, Aug. 1970.

Peter F. Donnelly, Federal Reserve Bank of New York: Public Utility Commissioner, Akron, Ohio.

Peter B. Kenen, Columbia University: Princeton University, June 1971.

Patricia Keough, Federal Reserve Bank of New York: New York University School of Economics.

Jacques Melitz, Tulane University.

Thomas F. Wilson, Federal Reserve Bank of New York: Butler University.

Miscellaneous

Arthur D. Lynn, Jr., Ohio State University: associate editor, *National Tax Journal.*

◆

The American Economic Review

The American Economic Review

CALCU
CENTRAL LIBRARY
UNIVERSITY

VOL. LXI, NO. 3 PART 2 JUNE, 1971

SUPPLEMENT

SURVEYS OF NATIONAL ECONOMIC POLICY ISSUES AND POLICY RESEARCH

1. MEXICAN ECONOMIC POLICY IN THE POST–WAR PERIOD: THE VIEWS OF MEXICAN ECONOMISTS—LEOPOLDO SOLÍS
2. YUGOSLAV ECONOMIC POLICY IN THE POST–WAR PERIOD: PROBLEMS, IDEAS, INSTITUTIONAL DEVELOPMENTS—BRANKO HORVAT

Price $1.00

Foreword

The two surveys of national economic policy issues and policy research published in this Supplement to *The American Economic Review* are the third in a series of studies of economics in foreign countries, commissioned by the Publications Committee of the American Economic Association and under the editorship of Harry G. Johnson. The previous series, edited by George W. Hildebrand, sought to acquaint Anglophone economists with the significant developments in economics since the Second World War in the national literatures of the major non–English–speaking countries. This series, by contrast, is directed at economic policy issues, and the controversies and research to which they have given rise, in countries selected for one or more of three reasons: the intrinsic interest of the policy issues, the relevance of the policy issues in the United States, and the interest of the countries themselves as areas of involvement of American foreign economic policy with which U.S. economists are likely to become concerned.

HARRY G. JOHNSON

previous sections, as well as on the activities of Mexican economists.

I. *Background*

The recent economic history of Mexico falls into two clearly defined stages. The first of these, characterized by development with inflation, covers the period from the end of World War II until 1956, with annual price increases of 11 percent, government deficits financed with currency issues, foreign exchange instability—the peso was devaluated in 1948 and again in 1954—and a rapid agricultural development. In fact, agricultural output had a higher rate of growth than that of the gross national product (6.2 percent) and brought about a considerable expansion in agricultural exports and foreign trade in general. In short, this was a period of externally-oriented growth.

The second stage, however, has been inwardly-oriented, based on substitution for imports of manufactured goods, the proportion of which decreased from 8 percent in the first period to 5 percent in the second. A relative stagnation has also become evident in the agricultural sector, the rate of growth of which fell below that of the gross national product, which has remained 6.2 percent. The exchange rate of the peso has been maintained at 12.50 to the dollar. Prices have only risen 1.9 percent on the average in recent years—less than in the United States, with which Mexico effects 70 percent of its foreign trade.

As a result of the diminished vitality of the agricultural sector, the rate of increase of agricultural exports has declined considerable, and in general receipts of foreign exchange have increased at a slower rate than expenditures abroad. This has produced a large deficit in the current account of the balance of payments and an increase in the foreign debt. The slowing down of investment in social overhead capital and in agricultural research and development, as well as the distortion of cost structure resulting from the effects of protectionism on the prices of industrial inputs, have had an adverse effect on the competitive position of agricultural exports and constitute a factor in the recent slowdown in growth of merchandise exports. On the other hand, the parity prices for goods destined for domestic consumption have been raised in order to stimulate their output. (The parity prices are set by the official regulating agency for agricultural commodities, *Compañia National de Subsistencias Populares.*)

One characteristic of the period of price stability has been the growth of domestic savings, which has played an important role in the accelerated expansion of the banking system and, in fact, in a general financial boom of unusual magnitude, brought about by factors such as transfers from the non-financial market to the institutional sector. The process has facilitated the placement of government securities by the central bank among the financial intermediaries—especially the non-monetary ones—so that government deficits no longer represent a mere creation of money by the central bank, as they did in the first period.

The Agricultural Sector

The agricultural sector has provided many of the outstanding features of the economy of the country. This sector absorbs 50 percent of the labor force and contributes less than 20 percent of the value added in the production process. These facts in themselves express the problems related to the distribution of income and the productivity of the labor force. The problems are much more pronounced when the dual nature of Mexican Agriculture is taken into account. There is a modern agricultural sector that can compete internationally in terms of efficiency and productivity and which generates about 65 percent of the agricultural output while employing only 36 percent of the agricultural labor force. The traditional sector, which continues to have a backward technology and a low productivity, employs the larger part of the agricultural labor force and produces ten percent of the value–added in agriculture.

Despite the problems mentioned, the agricultural sector has been strategic in the process of development. During the first stage of recent economic development 1946–1956, the average annual rate of growth of the gross national product, at constant prices, was 6.2 percent. This growth was characterized by the rapid development of the agricultural sector, which grew at a rate of 7.5 percent a year, that is, faster than the growth of domestic demand for agricultural commodities due to the increase in both population and real income. Thus, it was possible to improve the payments position of the country by substituting for imports of agricultural commodities and creating agricultural surpluses for the export markets. Foreign sales of agricultural products, especially those coming from the new irrigated areas, constituted the prime factor in the increase of exports, offsetting the decline in earnings on merchandise account due to the stagnation of mineral exports, which had been evident since the end of the Great Depression.

The rapid growth of the agricultural sector was fostered by the land reform movement and investment in public works to promote agriculture, which resulted in a better distribution of agricultural output and the creation of a commercial, modern and efficient agriculture. Moreover, research led to the application of new, superior techniques which lent a greater flexibility to agricultural supply.[1] Aside from the expansion of the area under cultivation, the irrigated areas increased more than proportionally, thus reducing the importance of random climatic conditions. Similarly, agricultural research led to the development of new and superior varieties of seeds and to the propagation of modern farming techniques, including the use of fertilizers and insecticides, especially in the irrigated zones. As a result, there was a more flexible supply, as well as increases in total agricultural output, in the yields per hectare, and in output per man. The dual nature of Mexican agriculture thus became more accentuated. On the one hand, a commercial agriculture emerged in the irrigated zones, with high productivity and a capacity to absorb technological change, which has increased the yields, uses modern inputs with a high import content and is not labor intensive. In contrast, there is the traditional subsistence agriculture, largely oriented to the production of corn, making intensive use of labor with low productivity, unreceptive to new techniques and new crops, and in which population growth represents an increasing pressure on the resources.

Since 1950, there has been a new trend in the allocation of public funds for investment. Works for the promotion of agriculture, which had had priority within public investment since the 1934–1940 Cardenas regime (when the redistribution of land gained momentum) became less important and, instead, emphasis was placed on public investment for industrial development. The effect of public works in agriculture, after a period of transition during which the large irrigation projects went into operation, consisted of an increase in the rate of growth of agricultural output. The new policy had the converse effect—a decline of public investment in agriculture, including education and research, that could not be offset with increased yields or the (now reduced) expansion of irrigated lands. In fact, the rate of growth of agricultural output decreased to 3.8 percent.

---

[1] The available figures show that between 1941 and 1951 there was an average annual increase of 2.9 percent in the area under cultivation and of 1.0 percent in the yields per hectare. The average increases between 1953 and 1962 were 1.3 percent in the area under cultivation and 1.9 percent in the yields.

The slower rate of growth of the agricultural sector was lower than the rate of increase in the domestic demand for its products. Consequently, between 1956 and 1967, there were no increases in agricultural surpluses for the export market.[2] Moreover, parity prices for commodities destined for the domestic market were raised by the regulating agency for agricultural prices, to the detriment of the relative price of export commodities, which did not rise in world markets. Besides, as a result of industrial protectionism, there were increases in the prices of the industrial inputs most frequently used in export products, such as cotton, and a subsequent loss in their profitability. Thus, both agricultural output and exports had low rates of growth. In the period 1955–1964, total exports expanded at an average annual rate of 4 percent, while those of agricultural products grew only 1.4 percent a year.

The pattern of behavior of Mexican agriculture poses serious problems to the economic development of the country. On the one hand, the modern agricultural sector has not increased output sufficiently to meet domestic demand, hence restricting the exports and therefore the import capacity of the country. On the other hand, the limited technological development of the traditional sector has made the absorption of the growing agricultural labor force increasingly difficult, reducing the levels of per capita incomes and bringing about a migration toward the urban centers, which in turn had generated cheap labor for industry and the service sector, exerted pressure on public utilities, and made it necessary to increase investment of social interest.

Industrial Development

Even prior to the drastic changes in agricultural production and in the yields per hectare, which permitted the growth of exports and secured the necessary foreign exchange for imports of machinery and equipment, industrial development has been the primary goal of economic development policy since World War II. Industrial firms were protected from international competition by means of tariffs and restrictive import permits. Tax incentives were created and, with the support of financial policy, an increasing volume of funds was channeled into industrial investment frequently by means of loans from official banks and international organizations at lower interest rates than those prevailing in the domestic market. Regulations were established so that private credit institutions would increase their industrial portfolios. Similarly, foreign firms, previously oriented toward raw material exports and public utilities, began to manufacture for the domestic market. The expansion of the industrial base strengthened the urban market for agricultural commodities in such a way that domestic demand supplemented exports as a stimulus to agricultural output.

These factors explain the 6.2 percent annual growth rate of manufacturing between 1940 and 1953—a rate almost equal to that of total output in the same period. In the years 1953–1965, the rate of growth rose to 8.3 percent a year, far above that of the gross national product in the same period. The development of the steel industry, the metal products manufacturing industry and the chemical industry spurred the growth of the whole sector. Yet some of the traditional industries—such as textiles, apparel and shoes—grew at rates similar to that of agricultural output. On the whole, the importance of industry was enhanced, as evidenced by the increase of its relative share of gross national product from 30.8 percent in 1950 to 35.7 percent in 1965. The relative share of manufacturing industry in the same aggregate rose from 20.8 percent to 25.7 percent. Nevertheless, this information in itself does not reveal the important internal structural changes within manufacturing. Other industrial activities developed in accordance with special conditions that affected their pattern of behavior.

---

[2] In the period 1956–1968, the rate of population growth was 3.4 percent a year and that of product per capita was 2.7 percent. Both factors combined to make the rate of growth of the demand for agricultural commodities about 5.5 percent.

The electric power and petroleum industries, stimulated by the government effort to strengthen the productive social overhead capital, expanded more or less at the same rate as manufacturing. Their average annual rates of growth in the first period (1940–1953) were 6.5 and 6.3 percent, respectively. In the second period (1953–1965), like manufacturing, they accelerated their rate of growth to 9.7 percent and 7.6 percent respectively.

The industrial evolution of Mexico between 1950 and 1965 was achieved by means of the accelerated production of capital goods and intermediate goods. In effect, ranking industries in accordance with the use of their products, there has been a decline in the importance within the industrial structure of those industries oriented toward the production of consumer goods. Consumer goods represented 72.2 percent of total manufactures in 1950 and 43.3 in 1965; the annual rates of growth for the fifteen-year period under consideration were 5.6 percent for consumer goods industries and 11.1 percent for producer goods industries. Industrial development in the same period strengthened the production of intermediate and capital goods. The dependence on imports of inputs necessary for the maintenance of equipment not produced domestically did not become more pronounced or more rigid.

The high level of tariff protection and quantitative restrictions on imports have meant high differentials of domestic over foreign prices, as well as high rates of effective protection which pose serious problems for industrial development. Similarly, although it is highly protected, industry has an unutilized capacity, uses capital-intensive processes, and has a limited absorption of labor.

Distribution of Income

In the last three decades, however, while Mexico has undergone a process of sustained economic development that has permitted output per capita to increase from 75 dollars in 1940 to 560 dollars in 1968, there have been indications that the process of development itself has engendered inequalities in the income received by the various segments of the population. The gains in output per capita must thus be qualified by adverse changes in the distribution of income.

In most countries, factor payments to labor represent a stable proportion of income. In Mexico, however, this proportion has been subject to wide fluctuations. Salaries and wages represented 29.2 percent of total income in 1940, but only 23.8 percent in 1950. In the period of growth with price stability, the trend was reversed, and the share of salaries and wages increased to almost the same proportion it had represented in 1940. The decline in the relative share of labor in comparison to that of capital throughout the decade of the 1940s can be explained by the rapid changes in prices and the greater use of labor during the war. This, together with the small possibility of increasing the stock of capital, meant that output growth could be obtained only through additional work shifts, thus increasing profits and the relative share of capital in the distribution.

The problem becomes patent when the data on income distribution by productive sectors and income brackets is taken into consideration. Agriculture, which is the activity that still absorbs the highest proportion of the labor force—50 percent—has an average level of income far below that of other sectors (less than 25 percent of the total in 1967). In addition, the distribution of income within the sector itself is very unequal, for 53 percent of the people engaged in agriculture receive only 23 percent of agricultural income. Inequality has been accentuated because commercial agriculture, with few units, grows more rapidly than the traditional, subsistence agricultural sector. In the period of rapid agricultural growth the relative (and even the absolute) position of low–income families self–employed in agriculture did not improve, while there was a rise in the income of landowners, agricultural entrepreneurs and farm hands of the advanced segment of agriculture in the irrigated areas.

Although initially the agrarian reform brought about a redistribution of income, its effect in absolute terms has been very small. On the other hand, public investment outlays on works for economic development, by making the modern sector of the economy more dynamic (especially in agriculture), have fostered a greater polarization of the economic situation without improving the lot of families in the lower income brackets, particularly in the agricultural sector.

In the industrial sector, the situation is much more equitable, for almost half of the people employed by it receive about 42 percent of the income. The mobility of rural labor toward the urban centers and its absorption by industry have not been sufficient to correct the inequitable distribution of income. Since the relative share of the agricultural sector within national income decreases faster than the proportion of the labor force engaged in agriculture, the distribution of income not only continues to be inequitable, but tends to become more so as economic development takes place.

In general, the role of the public sector in the distribution of income has not been significant. While it is difficult to evaluate the effect on the different strata of expenditures on social works, which have grown considerably, it can be assumed that they are not very important as a redistribution mechanism. Expenditures on education, which account for a large part of these outlays, have a greater effect on those who pursue their studies beyond elementary school, who, for the most part, do not come from the rural areas or the lower income groups. A similar situation is true of public health services, which primarily benefit the urban centers.

Some economic policy measures, such as the concession of tariff protection and quantitative controls on imports, tend to raise the price of consumer goods and thus distort the distribution of income. Moreover, these measures give rise to disparities between domestic and world prices, which reduce the real income of the consumers of protected goods. The associated transfers of real income favor profits and other returns to property, because the conditions of a surplus of labor imply a practically constant wage rate.

The Financial System

In the last ten years, the Mexican financial system has experienced a sustained and steady development, as a result of the growing volume of savings. In former years, the financial system was greatly limited in its operations by the uncertainty resulting from inflation and the low real rate of interest, for there were fixed limits on the nominal interest on bank liabilities and bonds. Voluntary domestic saving, measured by bank liabilities in national currency, remained fairly constant in relation to national income in the inflationary period. More recently, however, its rate of growth has surpassed that of the gross national product at current prices. The strengthening of the Mexican financial system has permitted a more efficient investment of savings. Also, its rapid growth has given the authorities a greater weight in the money market in terms of attracting and channeling funds through the financial system. Thus, some economic sectors that had not been considered attractive credit risks, despite being key sectors for development, can now obtain credit through the mediation of the monetary authorities, who have resorted to diverse mechanisms in order to facilitate their access to credit. The natural resistance of the commercial banking sector, which normally wants to invest in activities that constitute a low risk or of which it has previous knowledge, has thus been overcome.

It has therefore been possible also to transfer the voluntary savings of the community to the public sector and to finance government deficits without jeopardizing monetary stability. The funds transfered in this way represent up to two–fifths of the liabilities of the banking system. Consequently, the growth of the banking system itself has been conditioned by the annual increase in total liabilities. This type of financing does not have an inflationary effect, for it uses the real savings of the population and channels them toward non–monetary intermediaries. The liabilities of the Mexican banking system in domestic currency have increased at an annual rate of 18 percent over the past ten years. The growth rate, however, was slightly lower in 1967 and 1968, perhaps as a result of the decrease in the volume of savings outside the banking system, which during the latter years of economic stability constituted a significant part of the flow of funds toward the banking system.

Domestic savings channeled through private credit institutions have increased at an annual average rate of 21 percent since 1960; those channeled through the public institutions have done so at a rate of 22 percent. Within the liabilities of the private institutions, the growth of investment banks ("financieras") has been outstanding, for they have had a mean annual growth of 33 percent since 1960 and an even more marked increase in the past two years. This does not hold true for other private credit institutions, especially commercial banks, which have suffered a decline in their relative importance.

Although the instruments used to attract funds have become more diversified, their principal characteristic is still their liquidity, for even new financial certificates are convertible on demand, despite a fixed maturity date. The high liquidity of the system is one of its most vulnerable points, for in times of uncertainty the structure of holdings can be altered quickly to other real assets or to foreign currency, as occurred during the devaluation of 1954.

Public Finance

Mexican fiscal policy has had the following main objectives: the determination of a desirable composition of government spending in consumption and investment; the extension of services and assistance lent by the State to the community, spreading educational, medical and social welfare services to the population; the use of tax incentives to foster saving and private investment; making the tax burden more comprehensive by means of income taxes; and finally, the use of subsidies and tax exemptions to stimulate industrial development. With respect to the balance of payments, fiscal policy has consisted of creating incentives for industrialization by means of tariffs, exemptions and subsidies, in order to obtain substitution of domestically-produced goods for imports.

The main obstacles confronting Mexican fiscal policy have been: a persistent tendency toward budgetary disequilibrium and the problems of financing it in a non-inflationary way because of the weakness of the securities market; a concentration of the distribution of income and the problems of taxing the higher income groups on remuneration from various sources hard to discover; and the problem of taxing members of *ejidos* and cooperatives, which predominate in certain sectors, for political reasons.

The breakdown of the ordinary revenues of the Federal Government provides an idea of the tax structure of the country and its sources of funds. In 1967, total Government revenues amounted to 39.4 billion pesos—12 percent of gross national product. Of this total, income taxes represented 28.9 percent, while taxes on consumption accounted for 21.9 percent of the total, and taxes on income received from services and property of the State amounted to 4.2 percent of the total.

Another important source of Federal funds has been the public debt, although the rudimentary nature of the money and capital markets limits the volume of domestic debt. For this reason, especially since the slowdown of exports in the period of stability, foreign loans have been sought by the Government from international organizations and private institutions. In this way, domestic savings have been supplemented by foreign credits, with the result that there have been larger volumes of investment and equilibrium in the financing of development.

The recent income tax reforms represent an attempt to obtain larger volumes of the income of the community by consolidating the former cedular tax system —based on different rates depending on the income source—into an overall income tax rate. The reform of the tax system was begun in 1961. The taxable base was enlarged in order to include types of income which formerly had been exempt, such as returns on fixed-income securities, capital gains, and income from rentals of real estate. Similarly, a supplementary tax rate was created for accumulated income (as a precedent to the tax on total personal income), which offset to some extent the lack of progressiveness in the rates. In order to foster reinvestment, undistributed profits were exempted from taxes, provided that they were used for reinvestment. Another stimulus for investment was a scheme for accelerated depreciation.

The Federal Tax Bureau (Registro Federal de Causantes) was created in order to control those subject to taxation. Since the former system resulted in a low level of government revenue, and because it was inequitable (taxpayers within the same brackets paying different rates according to the source of their income), a change was made in 1965 to tax total income of both individuals and firms. The tax rate for firms with an income of more than 500,000 pesos is 42 percent. The level of taxable income was reduced for certain incomes received by credit institutions, insurance firms, and bond companies.

The recent trend of Mexican fiscal policy has been to reduce the dependence of the tax system on indirect taxes, which are regressive by nature, and to revise taxes with low yields which both make for high collection costs and induce a high degree of tax evasion.

However, there are still structural flaws. Studies show that it is still necessary to reform taxation on income and spending. Excise taxes include a number of items which have low yields due to high administration costs associated with the establishment of various levels of rates with no thought of integration; and often the rates are graded pyramidally, which is a drawback because of overlapping taxes. In 1966, further changes were made to integrate the budgets of government agencies and enterprises with that of the Federal Government. Similarly, all of the applications for credit of these institutions now have to be approved by the Ministry of Finance.

Public spending has been used to create the social overhead capital framework of the country. This was done initially by lending support to agricultural development and communications and later, in the recent period of stability, by doing the same for industrial development and works of social interest. In 1966, 42.9 percent of total public spending was allocated to economic development, 29.5 percent to investment and social welfare, 7.1 percent to the armed forces, 11 percent to administration costs, and 9.5 percent to servicing the public debt.

Balance of Payments

After the exchange instability and the boom in exports during the inflationary period had ebbed, there was a decline in the growth of exports, posing serious problems for the maintenance of the exchange rate and the growth of the national product. The problem was partially solved by means of financial policy, especially that concerned with maintaining the growth of the import capacity. This was achieved mainly through foreign credits, which, in turn, permitted a high rate of growth of the gross national product combined with price stability.

The deficit in the current account of the balance of payments has become more pronounced in recent years.[3] Between 1963 and 1968, expenditures abroad at current prices increased at an average annual rate of 10.6 percent, while foreign exchange earnings increased 6.7 percent. The result has been a larger deficit in current account, which in turn has meant an increase in foreign indebtedness. It is significant, however, that there was a change in the structure of expenditures abroad, for merchandise imports increased at a lower rate than overall expenditures and their relative importance declined, while that of expenditures by Mexican tourists abroad increased, as did the remittances on direct foreign investment and the interest payments on foreign debt of the public sector, which became twice as important.

Imports have shown a tendency to decrease as a proportion of the gross national product. This has been especially true in the period of price stability, of rapid industrial development, conditioned by import substitution.

High tariff protection has been a recent development, and very marked in the period of stability. The effect of tariffs is stronger than is generally thought. The tariff structure reveals low duties on raw materials, machinery and basic foodstuffs and high import duties on manufactured consumer goods.

In Mexico, the balance of payments is a secondary consideration in the formulation of trade policy. The strongest argument, since the decade of the fifties, has been the protection of incipient industries, as development became increasingly oriented toward the domestic market. During this period, the substitution for imports of durable consumer and intermediate goods gained momentum. But this process was very intensive as early as 1950.

The items which have contributed to the growth of foreign exchange earnings between 1963 and 1967 are income receipts from transactions along the border with the United States, tourism, agricultural commodities, manufactured goods and mineral products. Merchandise exports had the lowest rate of growth—4.2 percent. Foreign earnings from exports of services grew at a rate of 9.6 percent. The latter rate was abetted by tourism, which had an annual growth of 14.6 percent, making it the most dynamic item in the balance of payments, as well as by foreign exchange earnings border transactions with an annual growth of 7.5 percent. These trends have been reflected in the structure of income in current account. Merchandise exports fell to 50.2 percent of total exports in 1967, while income from tourism increased to 16.5 percent of the total.

---

[3] In the period between 1950 and 1968, with the exception of 1950 and 1955—years immediately following devaluations of the peso—the deficit on current account averaged 258.6 million dollars annually. In 1962, the deficit was 156.4 million dollars. The average for the last four years under consideration amounted to 446.1 million dollars.

As a result of the slow growth of merchandise exports, income on current account of the balance of payments available to pay for imports after subtracting payments to foreign factors (profits of foreign investment and servicing foreign credits) grew at an annual average rate of 6.1 percent between 1963 and 1967—less than the 8.5 percent rate of growth of imports.

II. *The Agricultural Sector*

In discussing policy related to the agricultural sector, it is necessary to distinguish between policy concerning land tenure and policy concerning agricultural production. The differentiation is chronological, as well. Land tenure policy was predominant during the period of intensive redistribution of land, starting during the Lazaro Cardenas regime (1934–1940). Later, interest gradually began to center on the problems of agricultural production stemming from agricultural development and the new pattern of land tenure.

Land Tenure

What were the results of the agrarian reform? Under what conditions is the land tilled in Mexico? Agrarian reform changed the structure of the ownership of land and, consequently, wrought changes in the rural social structure itself. Its principal goal was social justice—the distribution of land among the peasants who work it. For this very reason, it lacked an economic criterion as a guideline for the distribution of land to establish productive units of an efficient size. Because of this, there is still much discussion about the optimal size of plots. However, according to E. Flores (1967), it involves a good deal of ideological discussion and few reliable studies of the economic feasibilities.

With the land reform came the *ejido* system,[4] giving a communal solution to the agrarian problem. This system was created mainly on the basis of the lands expropriated from the former *haciendas* (large land holdings), which, by the same token, brought about the disintegration of the latifundia. The disappearance of the latifundia had two immediate consequences: the destruction of a great number of productive units and a subsequent decline in the productivity of others, and the divorce of economic and political power from the traditional rule of class (R. Stavenhagen, 1966, pp. 468–469).

A form of small private ownership of the land emerged together with the *ejido* system. In 1960, barely 25 percent of the agricultural labor force was made up of *ejidatarios,* which reveals that the majority of the rural population lives under a system of private enterprise. There are also owners of medium sized properties and family farms, lying somewhere in between the small and the large landowners.

Most opinions concur that the *ejido* is in crisis or stagnant and is not an institution than can provide a satisfactory solution to the agrarian problem. Stavenhagen points out that the *ejido* system of land tenure has created several problems as the result of communal production under the conditions of a capitalist economy. For example, the collective *ejidos* have been infiltrated by interests and political forces which have thwarted them. Within the *ejido* system, the main problems are a) the smallness of the individual plots or *minifundios,* b) the uncertainty of tenure, and c) the scarcity of credits for the *ejidatarios*. These problems, which stem from the smallness of individual plots, hinder the formation of capital as well as an increase in productivity, and maintain the technological level of the rural sector stagnant. R. Fernandez y Fernandez (1957) points out some of the symptoms of the situation—the sale of plots to strangers, the marked government paternalism, the rigidity of the institution (or the fact that it merely represents a subsidiary form of land holding, which is not very efficient).

---

[4] On this *ejido* system the land is owned by peasants, who can work it and bequeath it, but can neither sell it nor rent it.

A problem to which other authors refer is that of the lack of flexibility in the performance of the *ejido*. The more efficient *ejidatarios* cannot enlarge their holdings of land—which, by law, they cannot rent—nor can they displace the less efficient. The problem has become more acute because of demographic growth and the lack of new lands in areas where the agrarian reform already has been effected. Nevertheless, there are many instances of *ejidatarios* who lease their land illegally in order to become salaried workers.

But how did these conditions arise? The consensus of opinion points to the very mechanism of the agrarian reform and the organization with which the *ejidos* were actually endowed as the principal factors. Fernandez, M. Hinojosa Ortiz (1961), and Stavenhagen make the observation that in the distribution of land, recipients were not screened nor were they organized technically or institutionally. The corps of technicians in charge of implementing distribution was insufficient, both in number and in qualifications, and generally lived in the cities. On the other hand, there have been numerous shortcomings in its operation, and its organization on the basis of assemblies and commissaryships has only been an additional problem. There has also been a dispersion of the jurisdiction of the authorities which has led to confusion among farmers (Hinojosa). In conclusion, the *ejido* was a system that was conducive to the substitution of labor for capital (M. A. Duran, 1968), in an environment which is still overpopulated (Fernandez, 1957). As Stavenhagen remarks, besides the fact that the majority of the *ejidatarios* were given extremely small plots of inferior quality, the lower productivity of the *ejido* can also be attributed to the irrigation policy, which accorded greater benefits to the owners of private plots. The *ejido* has lost its initial momentum and, at present, it is being thwarted by the private agricultural sector (Fernandez, 1957), which has the most productive form of landholding (Stavenhagen).

The proposed solutions to the problem contain many viewpoints in common. In this respect, Fernandez y Fernandez (1957) speaks for many authors. He points out the need to select the new *ejidatarios* who will receive land to be distributed; to determine which *ejidos,* and in what fields, will adopt the cooperative form (with regard to the existing cooperatives there are some discrepancies). For some, this form of organization is the best course to follow and its radius of action should be enlarged (Duran, 1961; Fernandez, 1954); for others it is merely an abstruse imitation with a complex structure (H. Flores dela Peña and A. Ferrer, 1951). Fernandez also stresses the need to increase the size of the plots held by every *ejidatario* to the size of family farm plots; to make it possible for the *ejidatario* to purchase land; to chastise the *ejidatario* who does not work his plot; to organize them, seeking a structure to foster internal cooperation; to maintain a system of equal plot sizes by regions, since it is practically the only feasible one (however, others differ on this and would prefer that size depend on the economic conditions of the land); and to insist that government intervention be limited to the administration of justice and technical assistance.

Within the private sector of agriculture, the principal problem is the small private plot and the conditions around it, which are getting worse. The majority of the owners of small plots live on a subsistence level. Like their *ejidatario* peers, they are characterized by low productivity and a limited technical level, besides which they are still in the hands of local money-lenders. Many of them have to look for work as migrant farm laborers in the United States, or as temporary laborers on large farms, or migrate to cities. On the other hand, commercial farmers are very efficient and with highly mechanized enterprises. In the small plots and the *ejidos,* there is a surplus of labor which does not favor the formation of capital. A great number of them produce for home consumption. The small plot hinders technical progress. The incidence of the role of technological change is reflected in that part of output earmarked for home consumption (F. Rosenzweig Hernandez, 1963). This is the reason why greater productivity can be obtained on larger properties and agricultural progress is more marked in the areas under irrigation (Flores dela Peña, 1953; G. Lira Porragas, 1964).

Similarly, regarding the commercialization of agricultural products, there are monopolistic conditions conducive to price-fixing that hamper the formation of capital in agricultural units. It is frequently pointed out that the middlemen make the largest profits, for they buy from small, unorganized farmers—to whom they often have lent money—and later sell as monopolist to the regional market (Rosenzweig, 1963; Stavenhagen, 1966).

For Rosenzweig as well as for Stavenhagen, the small plots constitute the main problem of the agricultural sector today, above all because they impede the formation of capital for lack of stimuli. Moreover, the situation is taking a turn for the worse. No coherent policy has been formulated to solve the problem and it cannot evolve "naturally," due to the structure of the markets and the lack of technical education.

The solution for the small private and *ejido* plots, as seen by Fernandez (1954) and Rosenzweig (1963), is to integrate the fractionized property into plots of an economically efficient size. At the same time, the small farmers, who constitute a weak and scattered group, must count on greater government protection. Another writer, Flores de la Peña (1954), argues that if the system of extensive exploitation is continued, there will also be the drawback of not solving the population problem. Because of this, small farms of the European type—intensively worked—should be established as a measure to solve the problem of land tenure as well as that of overpopulation.

One cannot overlook the advance in the productivity of agriculture. The social overhead capital works executed by the government permitted substantial progress in various areas, especially in the northern part of the country. Similarly, new means of communications, namely highways, allowed easy access to markets.

There are two positions regarding what is almost always called the *latifundium,* that is, the commercial exploitation of relatively large extensions of land. One of them, upheld by Hinojosa (1961), considers that the Agrarian Reform put an end to the *latifundium* in Mexico. The other, sustained by Flores de la Peña, Fernandez and Stavenhagen, among others, is that the existence of new *latifundia*—modern, commercial agriculture on large properties—is a fact in the country today. According to Flores de la Peña, Mexico after 30 years of Agrarian Reform, is a country of large landholders. In Stavenhagen's opinion, this situation is recent and for that reason he speaks of neo-latifundia. He claims that the new landowners are often very efficient, highly mechanized agricultural enterprises and that they represent the negation of the very ideals of the Agrarian Reform (1966). Flores de la Peña (1954) and Lira (1964), point out that the concentration of property can be found particularly in the areas under irrigation, and they are in agreement with Stavenhagen in attributing the prevalent form of *latifundia,* of recent origin, to the way land was distributed in newly-irrigated regions. In any case, all agree on the need to abolish the *latifundium.* Fernandez, (1960) favors forbidding it by law and attacking clandestine or de facto *latifundia.*

Underemployment and Colonization

The phenomenon of rural underemployment is universally recognized. It is also a factor that plays an important role in recommendations for development policy, as will be shown later. Opinions on this topic supplement each other. For Flores (1964, Ch. XX), the principal problem arising from this situation is the waste of human resources. One of the reasons for disguised unemployment is the traditional outlook of the population and its concentration in certain regions. Flores de la Peña feels that rural underemployment is a problem that has yet to be solved and because of this proposes the establishment of small farms with an intensive use of labor as the predominant mode of agricultural production.

A number of solutions have been proffered with respect to the problem of underemployment. The main one is a measure that would be feasible only in the long run: the absorption of the surplus rural population by secondary and tertiary activities. This, however, would depend on the growth of these activities. Flores (1964), Flores de la Peña (1958), and Rosenzweig (1963), agree with this viewpoint, but there are others. For example, according to Flores de la Peña (1954), the rural population should be frozen at six million people, as of 1960, who would be guarenteed work all year round. He also makes other recommendations, such as the displacement of groups of farmers towards the northern part of the country or the Gulf region, which contain the most productive lands. Finally, he is of the opinion that underemployment will not be solved by the modernization and mechanization of agriculture nor by the distribution of lands (1958, p. 378).

Colonization has been carried out very slowly (Fernandez, 1960) and is increasingly costly, since the best lands have already been distributed (Rosenzweig, 1963). For Stavenhagen, colonization has been important in recent years: because of it, the increasing importance of internal migration in the country is due to factors which are not directly related to agrarian reform (1966, p. 468). Flores de la Peña considers colonization preferable to the distribution of lands under cultivation, for reasons of capital formation and the improvement of overall efficiency. Fernandez advises the acceleration of this process in view of the overpopulation on the *ejidos.*

Agricultural Policy

Capital Formation. On the whole, there has been a limited capitalization of the private segment of the rural sector. There is general agreement on this, even though there are differences of opinion regarding its causes. Rather than being antagonistic, however, the opinions tend to supplement one another. Fernandez (1960) believes that the reason lies in the lack of incentives, emanating, as shown above, from the pattern of behavior of the *ejidos*. On the other hand, Rosenzweig holds that the reason can be found fundamentally in the monopsonistic practices of the markets, as well as the deterioration and forsaken condition of capital equipment on some of the land (1963). To this, other experts add the insufficiency of credit.

The evaluation of public investment is different. The works executed by the government, such as communications and irrigation projects, are considered essential factors in the determination of the increase in productivity experienced in recent decades. Thus, Flores (1959) influenced by J. Schumpeter, speaks of the "strategic innovation" of these projects, in the sense of the effect that they had on the entire productive structure and its process of transformation. At the same time, they fostered the relocation of agriculture, displacing it toward the northern and northwestern parts of the country, as well as stimulating migration toward new farming areas.

In view of the foregoing, the solutions proffered are obvious: all insist on more intensive investment, private as well as public. For some, like Duran, Flores and Rosenzweig, planning must be introduced into agriculture. Others consider that the irrigation and communications policy must be continued, while the credit facilities and technical assistance granted by the government are expanded (Flores de la Peña, 1958). Rosenzweig adds that there should be a more extensive use of chemical inputs in agriculture (1963). Finally, Hinojosa manifests some dissention by suggesting that the imbalance between large and small irrigation be corrected (1961). In this, he sides with Stavenhagen, who states that the irrigation works have basically benefitted not the ejidatarios and small farmers, but the large landowners (S. de la Peña, 1963).

Commodities and Markets. Flores links the relocation of crops with the importance acquired by cotton, coffee and henequen, which were organized as commercial agricultural production oriented toward foreign markets. Yet, another phenomenon has emerged: these commodities are being displaced by fruit and vegetable crops because of the higher productivity of the latter two (1959). Regarding wheat, there are two contradictory opinions. Flores maintains that this commodity has not reached the position attained by the others, despite a favorable domestic price and credit policy. However, Lira (1964) believes that it is the only commodity within Mexican agriculture that has good yields.

The opinions about markets can be separated into those pertaining to foreign markets and those referring to domestic markets. Most experts have doubts about the stability of the foreign markets. According to Duran, their instability is always latent (1961). Similarly, Lira affirms that difficulties are still encountered in placing the output of the northwestern region in foreign markets (1964). There is, furthermore, agreement on the flagging vitality of the domestic market, probably a reflection of Engel's law. Fernandez (1954) and Flores de la Peña (1954) attribute the lag in the consumption of the agricultural masses to the deterioration of their terms of trade with the industrial sector. Moreover, Flores de la Peña refers to the lack of an economic policy on basic commodities. On the other hand, Rosenzweig (1963) and Flores de la Peña (1954), point out the prevalence of monopsonistic situations in the domestic market. The first of these authors further indicates that the low level of domestic consumption is a result of low productivity and that agricultural demand is weak and easily satisfied.

All of these authors are in favor of enlarging the domestic market. Rosenzweig offers some concrete solutions with regard to the objectives for economic policy, such as the organization of the market and producers, the elimination of middlemen, improving the commercialization of agriculture and abolishing monopolistic situations. Fernandez (1954) stated in the early 1950s that the State should intervene marginally in fixing prices by means of regulating inventories.

Technology and Credit. The paltry investment in agriculture is related to the low level of technology (Flores, Rosenzweig, Duran and P. Padilla). Yet, there have been regional differences in the use of technology and the gains have not reached all areas evenly, according to Hinojosa (1961).

The problem of credit has been dealt with extensively, so that it is possible to infer that it manifests itself as a phenomenon of real importance in the rural sector, and that it is closely related to the insufficient investment and the market situation. All of these authors agree that investment is insufficient in terms of the "needs" of producers, but there are different opinions about the relationship between cause and effect. According to some (Fernandez, Duran, Flores and Flores de la Peña), the lack of credit is a determining factor in the low productivity of the rural sector, since it is insufficient, poorly timed, and centralized. Others (Rosenzweig and especially Stavenhagen) consider that the lack of credit is a result of the mode of agricultural production—namely small farms and the market situation—which makes the credit risk involved very high, so that the private banking system does not meet the needs of the sector. Finally, it should be mentioned that credit to *ejidos,* in practice, only benefits a small proportion of the *ejidatarios* (Stavenhagen, 1966) and in many instances, there has been malversation on the part of those in charge of *ejido* credit funds (Hinojosa, 1961). Those authors who have dealt with the subject point out the need for improvement. Fernandez and Flores specifically refer to the need to dencentralize the credit system. Lira emphasizes that it is necessary to extend the terms of credit and to allocate it to annual crops and fruticulture. Rosenzweig does not believe in half-way measures—it is true that improvement is necessary, but it should be conditioned by other measures to increase productivity (1963).

General Analysis of Agriculture

A general appraisal of the agricultural sector embraces two outstanding aspects —the result of the agrarian reform and the role of agriculture within the context of economic growth.

The agrarian reform had positive and negative aspects. According to Flores (1959), its most important contribution was essentially the redistribution of wealth, income, and power. With this, it gave the initial impetus to economic development, propitiated the mobility of factors of production, and altered the pattern of land use. Moreover, it spread its influence to other economic sectors. Fernandez (1954), and Rosenzweig (1963), consider it the reason for public investment in agriculture, that is, in social overhead capital projects such as irrigation, communications, and electricity, which later helped to increase productivity. At the same time, it was not without flaws, stemming from the absence of a well-defined program and its fundamentally political motivation, as well as economic limitations and human shortcomings (Rosenzweig, 1963; Stavenhagen, 1966). This has led Stavenhagen to the conclusion that the reform has not yet been completed, for it has only attained its goal of redistributing the land and has inadvertently permitted the emergence of large landowners, with an efficient production closely tied to commercial agriculture.

The role of agriculture within economic growth has not been less important. Flores de la Peña (1954), considers that it has been essential in maintaining an external equilibrium—an opinion shared by other economists. Nevertheless, the sector has not been favored, as the same author points out, for development has meant a decrease in the real income of the rural sectors. The contraction of the effective demand for agricultural output has only accentuated this phenemenon (Flores de la Peña, 1958). Other authors find the limitations on agricultural growth in the behavior of the factors of production—the scarcity of capital and the surplus of labor (Padilla and Rosenzweig). Finally, Lira (1964), leaning toward a concept to be examined later, feels that the disequilibria of the sector are of a structural nature and that their solution requires economic planning in agriculture.

How can the ideas examined thus far be summarized? What stands out is that, despite its real achievements and impact on the present economic expansion, the agrarian reform has left some situations that require change, so that the agricultural productive system can reach a level of efficiency that would be compatible with the needs of the agricultural masses. Demographic pressure in the countryside is substantial, which merely represents an additional element that impedes the formulation of solutions.

One final appraisal is necessary: there is a scarcity of works that analyze livestock breeding and the recent stagnation in the production and exports of agricultural commodities. In speaking of the rural question, almost all of the authors are referring to topics related to agriculture, while the livestock aspect receives only marginal attention, such as recommending its increase in new areas or fostering consumption in the domestic market. It is well known that there is a small consumption of animal proteins and a high income-elasticity in the demand for these products.

III. *The Industrial Sector*

Background

In the past twenty years, industrialization has been interpreted as the very essence of economic development in Mexico. In the vainglory of nationalism, industrial self-sufficiency represents both the economic independence of the country and the foundation of its progress. Consequently, despite the importance allotted to the agrarian reform, it is undoubtedly true that since 1940, industrialization has constituted the main concern of economic policy. And until very recently, it has been unanimously supported by economists. There is no opposition, only a divergence of opinion regarding the pattern of the process and the sectors which should be given priority. Moreover, in the latter part of the 1950s, and even today, once the industrial base had been established, one topic has been subject to frequent discussion—foreign investment, which has experienced radical change in structure. Indirect investment has increased while direct investment has changed from being a supplier of raw materials and public utilities to being concerned with manufacturing for the growing Mexican market.

A second topic of importance that began to attract the interest of many economists in the sixties has been that of protectionism for the promotion of industry.

The interest was born from the participation of Mexico in the Latin American Free Trade Association and the plans for economic integration. The efficiency of the Mexican industrial apparatus, as well as its competitive position in world markets, became a matter of concern. Another factor, to be discussed later, which also contributed to this, was the stagnation of exports of primary products toward the end of the 1950s. Since then, there has been a great preoccupation with increasing exports of manufactured goods in order to raise the rate of growth of exports, in view of the limited prospects for agricultural exports.

Foreign Investment

The traditional type of foreign investment, oriented toward the exploitation of natural resources and public utilities, which had become firmly entrenched in Mexico by the turn of the century, practically disappeared after the oil expropriation (1938).

In recent years, direct foreign investment has established itself basically in the manufacturing and service industries, oriented toward the domestic market, abetted by the policy of import substitution and high profits. This has also been the result of the Mexican economic policy of reserving certain basic sectors of the national economy for the State and leaving others to the private sector. Moreover, there is a requirement of a majority of Mexican capital in those sectors considered as intimately connected with basic industries.

With regard to the composition of direct foreign investment, despite the recent diversification of the countries of origin, that of the United States accounts for 70 percent of the total. As far as the balance of payments is concerned, there has been a trend toward a growth of remittances abroad in relation to the influx of new capital. The observation, resulting from a comparison of a stock of capital with part of the income derived from its use, is supplemented by another element—the payments to foreign capital for such items as interest, royalties, and technical assistance grow at a faster pace than the profits. Within the Mexican economy, an increase has been experienced also in the reinvestment of profits of foreign companies in relation to new foreign investment.

Finally, since 1942, indirect foreign investment, in the form of international loans for development, has become increasingly important, especially for the financing of social overhead capital projects executed by the public sector.

The opinions of some Mexican economists with regard to the results of foreign investment and the measures they recommend for its treatment are presented below. However, it should be remembered that the appraisal of foreign investment in Mexico is often associated with ideological positions or the defense of vested interests. One reason lies in the lack of objective studies on the subject. The most serious and impartial works are either descriptive or make a very general analysis.

The first opinions to be presented are those of the experts who are in favor of the broadest possible scope for foreign investment, followed by those which express the viewpoints of the entrepreneurial and banking interest. A work by G. R. Velasco (1955) illustrates this current of thought. It states that foreign investment is beneficial, for it introduces new technology and creates new sources of employment for local labor. The author believes that it has been insufficient mainly for two reasons. First of all, there are fewer sources of capital today than before the war (1955). And secondly, there is a diffidence about investing in the Latin American countries because of their instability. Because of the requirements of economic growth and the scarcity of capital, Mexico needs foreign investment—it is the only thing that can raise productivity and accelerate development. For that reason, Velasco recommends an "open-door" policy. And, upholding a liberal economic concept patterned after the nineteenth century model, he considers direct foreign investment as the most convenient, for it guarantees the freedom of private enterprise and development. On the other hand, indirect foreign investment is unstable, has a political character, increases the influence of government in the economy, and leads to a less efficient use of resources (p. 18).

A diametrically opposite viewpoint is held by the Marxist left. A work by J. L. Ceceña (1965) can be considered representative of this current of thought. His approach is two-fold, as he repudiates two forces, international imperialism from without and the ruling oligarchy within a capitalist, underdeveloped economy. Foreign investment is the bond that ties and maintains the power of both of these oppressive groups. Those who uphold this position also manifest a very intense nationalism, and, on occasion, seem to be unaware of the structural change experienced in the pattern of foreign investment in Mexico that has taken place over the past three decades (F. Carmona, 1963). The main arguments are that foreign investment is an obstacle to development because it makes the economy tend toward disequilibrium and make it vulnerable to changes abroad. Moreover, it is conducive to a dangerous dependence on the United States, both in the markets for Mexican goods and as a source of supplies for national industry.

The nationalist intent makes this group of economists argue that economic independence is identified with industrial power, and foreign investment is harmful because it introduces disadvantages to the industrialization process. For one thing, the acquisition of modern techniques of production and business administration is limited, for the foreign subsidiaries operate with secondhand processes, which reduce industrial efficiency. It also absorbs smaller national firms or subjects them to the market conditions established by foreign firms. Furthermore, it gives place to a marked economic concentration which permits the use of monopolistic practices that only bring about domestic price increases. It is stated that of the 400 largest firms operating in Mexico, foreign enterprises make 36 percent of total sales. With supplementary data, it can be deduced that there is a "dual" economy, in which more than half of the firms are operated by head offices abroad. Thus, economic planning and industrial development are impeded, for the long-sought integration of national industry is only achieved by the head offices and their subsidiaries.

The leftist economists also find that banking makes for foreign dependence for reasons tied to foreign investment. Commercial banking is dominated by strong foreign interests and is greatly concentrated. Centering around the most important banks, a number of enterprises have been established that supplement, in financial terms, a complete system of economic power. Thus, banking is an instrument used by foreign monopolies to make use of domestic savings. This, in turn, has three consequences: easy access to the liquid resources of the country, a greater influence on business, and ties with the interests of national capitalists (Ceceña, 1965, p. 292). Moreover, the effect on the balance of payments represents a disinvestment for the country, because of excessive outflow of profits in the form of foreign exchange. A number of corrective measures have been proposed. To begin with, the leftist economists assert that economic policy must have two well-defined goals. On an international level, the country should pursue an independent policy trying to establish closer relations with Latin America. On the national level, the policy to be followed should be revolutionary and nationalistic, and should lend support to Mexican entrepreneurs by means of credit, technical aid, etc. Consequential to their belief in state intervention in all of the economy, however, they demand a policy applicable to foreign investment, with special norms, since at present the interests of the country are not defended. The great monopolies must be combatted by means of legislation and a planned regulation.

The third viewpoint, eclectic and more pragmatic, is held by a group of economists who, except for transitory differences, can be considered to follow the present official position. In general, they admit that foreign investment has complemented domestic savings, for it has contributed to doubling the per capita gross national income since 1939.

If the official statements on the subject of foreign investment were analyzed, it would be found that there are definite policy norms. Foreign investment must not displace national capital, which will have exclusivity in basic activities and majority control in those branches connected with the basic industries. It must be subject to the laws of the country. The preferred priority is: inter-government loans, international loans and direct foreign investment. Of the last, those that would be best received are the ones bringing technological innovations.

According to Alfredo Navarrete, who is an exponent of this position, there are other supplementary criteria. Direct foreign investment, which is now concentrated in the manufacturing industry, is geared to supplying the local market. This change in its orientation has made for a greater rigidity in the balance of payments, because formerly, with the traditional pattern of foreign investment, the remittances of earnings moved parallel to export income (Navarrete, 1966). The association of national and foreign capitals is an efficient procedure to attain technological progress. Local investors participate actively in the determination of profitable fields, while foreign capital has the technical responsibility. Mexican economic policy is in favor of the fusion of capitals with the object of increasing domestic capital formation. Actually, this has been the result not only of policy but also of the economic maturity of the country, due to the existence of an entrepreneurial as well as an investing class that favors this process (Navarrete, 1966). In the financial sense, foreign investment has not altered the debtor status of Mexico, although it has a smaller debtor position today than in 1939, because the increase in the foreign debt has been invested in productive works that enhance the payments capacity of the nation (Navarrete, 1958). The inflow of capitals is advantageous to both the recipient and the investing country. The effects of development loans on the balance of payments depends on the productivity of the investment, the rate of interest due on the loan and the length of time for which the loan is extended. In the case of direct foreign investments, these effects depend on the rate of profits and the reinvestment policies of the investors. The recommendations of Navarrete for the appraisal of foreign investment are based on the following criteria: a) complementary (in terms of domestic saving); b) flexibility (in the composition of investments); and c) equilibrium (between domestic and foreign saving and between direct and indirect investment). He points out that Mexico needs capital and for that reason it should also take advantage of the institutional saving of the insurance companies of the developed countries as another means of increasing the resources available for investment. But it is preferable that public investment be financed with domestic saving, generated by the voluntary savings of firms and individuals (Navarrete, 1958).

It would be interesting to mention two additional points regarding foreign investment. The first of these is the difference in the standpoints of the two private industrial sectors, the Confederation of Industrial Chambers (CONCAMIN), which represents big business, and the National Chamber of Manufacturing Industry (CNIT), representing medium and small industrial firms. The second point is related to an essay by M. S. Wionczek (1968) regarding technology, foreign investment and derived problems.

The CONCAMIN standpoint, as expressed by Campillo (1966), is fairly close to the official position. Foreign investment is a useful supplement to domestic saving in fostering development. It should not displace national capital and, preferably, a partnership of the two should be formed. It should be oriented to activities of social interest, providing jobs for Mexican technicians and personnel, as well as introducing new techniques. To judge it, it is necessary not only to consider reinvestment and the remittance of earnings but also to take into account the substitution for imports, the increase in exports and the multiplier effect of (foreign) investment on the economy in general (Campillo, 1966, p. 13). He does not believe that it would be beneficial to have a law to regulate foreign investment, but he does establish the need for a foreign investment policy that would unify the criteria of the various government agencies.

The appraisal made by the CNIT is just the opposite. Its exponent, R. A. Ollervides, asserts that domestic saving has been the real lever for development, not foreign investment (1966, p. 491). The latter could serve as a support if there were legislation to regulate it. Domestic saving is not insufficient, as bankers claim, but it is poorly channeled because of the excessive amounts of credit granted to commerce to the detriment of industrial activities. It is the banking system which is insufficient. There is no need to divert attention to other areas while defending vested interests. Finally, he argues, foreign investment has accentuated the exchange disequilibria and disinvestment of the country. Referring to a statement of the bankers, seconded by CONCAMIN, he maintains that it is difficult to assess the beneficial effects of foreign investment on the current account of the balance of payments. Moreover, it "generates additional imports that not only mean additional drawings of foreign exchange . . . but through clever devices used to import foreign machinery, equipment and material . . . deprive domestic activities of demand necessary for their development" (Ollervides, 1966, p. 490). The same author adds that the effect of foreign investment upon the ". . . so-called current account of the balance of payments is indirect and its influence on it very difficult to measure" (p. 490). International loans have other disadvantages, such as tied credits, which represent a subsidy for the creditor nation and impede industrialization on a national scale (Ollervides, 1966). For this reason he demands state intervention, so that, on the one hand, the basic and related industries be nationalized or Mexican-owned, and on the other hand, foreign investment not be completely abolished, but strictly regulated by means of appropriate legislation, since the present standards are not adequate for the interests of the country.

How can the difference in attitudes of the private industrial sectors be explained? Some grounds could be found in the conflict between manufacturers and importers that existed toward the end of World War II. Nevertheless, the present friction between small and medium producer and big business and banking is the outgrowth of the process of industrialization itself. Two disadvantages for the former that do not exist for the latter can be deduced from the statements of Ollervides —the scarcity of credit and the difficult access to external sources of intermediate goods. In substance, they seem to feel that industrial development is leaving them behind, and at the mercy of the big firms, because of their position not only in the market for goods, but also in that for factors of production. For this reason, they turn to the protection of the State, asking for its intervention from a standpoint reminiscent of that of the Marxist left, wherein the defense of interests and nationalist elements are combined with market arguments.

A second issue of interest is that of the technological gap and foreign investment. Wionczek (1967) begins by pointing out that within Mexican industry, foreign investment has shifted toward the most profitable and dynamic sectors. This has increased its profits and reinvestment, which could jeopardize the balance of payments position were the process to be reversed (no new investment and the repatriation of profits).

The increasing dependence of the country on foreign technology implies a new cause for friction, in view of the inequality of strength of large corporations and Mexican entrepreneurs (see the position of the CNIT). This is disturbing because "second-hand technology is received at an exaggerated price" and because technological dependence is associated with the introduction of new techniques by means of private foreign investment (Wionczek, 1967). Yet, due to "the quasi monopolistic conditions of the firms, it would be naive to ask for the liberalization of the conditions of the transference of new technology and its divorce from the investment of "capital" (p. 985).

Industrial Protectionism

The process of industrialization has been closely linked to protectionism since its beginnings in the nineteenth century. It was not until the Second World War, however, that protectionism became the basic instrument of industrialization. The larger demand during the war, resulting from the increase in exports and the budget deficit caused by the increased spending on public works, could not be satisfied entirely, regardless of the much greater utilization of the installed capacity of firms. Thus, the inflationary process became more marked. Moreover, there was some negligence regarding the standards of quality of manufactured goods.

Toward the end of 1945, the Law for the Promotion of Manufacturing Industry (Ley de Fomento de Industrias de Transformación) was enacted. It defined new and necessary industries and provided a stimulus for newly-formed basic manufacturing enterprises. The balance of payments problems, resulting from the disequilibrium of prices at home with those abroad, as well as from the deferred war-time demand, made it necessary in 1947 to establish a system of import permits, especially for luxury goods. R. Izquierdo (1964) states that at the end of the 1940s there was great enthusiasm for protection. The government championed industrialization and appeared to have found the formula for sustained and accelerated growth. Moreover, requests for protection from foreign competition were willingly considered, without careful attention to the type of product or the extent to which national materials were employed in its manufacture (p. 268).

The decade of the 1950s was characterized by two different phases of protectionism, separated by the devaluation of 1954 and the price stability attained subsequently (1957). After the devaluation of the peso in 1954, two goals of economic policy became apparent: to avoid domestic price increases and to provide incentives for foreign investment. The latter was justified in that protectionism had already altered the structure of imports, leading to the predominance of imports of capital goods. This, in turn, represented a greater inelasticity of demand for imports, which made manifest the dependence of industry on supplies from abroad. After the attainment of price stability, protectionist policy, based on the instrument of import permits, came to have as its fundamental objective domestic industrial integration. Import substitution and the consumption of domestically-produced goods was stimulated thereby. The system of licensing imports proved to be very effective because of its flexibility.

Opinion regarding the protectionist policy is a consensus that it has been necessary and beneficial for the goal of industrial development. However, its usefulness as an isolated tool is coming to an end and it should be reformulated in view of the present conditions of development of the country. Conditions are no longer such that by merely limiting foreign competition there can be a promotion of economic growth and industrial diversification. In the future, substitution for imports will be effective only in conjunction with the integration of national industry (Izquierdo, 1964).

In general, also according to Izquierdo, the Mexican experience shows that the formula of industrialization through protection is not self-perpetuating. Furthermore, the government has almost always given favorable replies to requests for protection, it has done so without due consideration to the type of product or its proportion of imported inputs, and without demanding the fulfillment of progressive integration programs. What might be called the "natural" theory of import replacement was widely accepted (Izquierdo, 1964, chapter 4). On the other hand, P. Garcia Reynoso stresses that industralization was attained at the expense of other sectors, e.g., agriculture, as well as through the sacrifice of the consumer. In relation to its effects, Garcia Reynoso also mentions some distortions wrought by protectionism on the Mexican economy. He specifically points out three outstanding problems: A) High costs, resulting from an indiscriminate protectionism, have rendered manufacturing industry inefficient. This particular trait limits demand in the domestic market and reduces the competitive position in foreign markets (1968, p. 963). (J. Saenz [1957] had already pointed this out in 1957, when refuting the ECLA theory on external disequilibrium, to which we will refer below). B) Geographic concentration of industrial activities. C) The limited impact of industrialization on total employment. Saenz criticizes the idea that industrialization per se overcomes unemployment, because it overlooks demographic growth and the intensity of the technological revolution. He indicates that this problem has become more acute lately because the technology that Mexico absorbs comes from highly developed countries where labor is the scarce factor and uses capital-and technology-intensive methods of production. Therefore, as long as Mexico improves her industrial structure and develops basic and intermediate industries, the marginal increments in employment are shrinking relative to the rapid growth of the labor force (Garcia Reynoso, 1962, p. 964).

The industrial development policy has not been free of problems. Izquierdo indicates that in terms of the process followed in its implementation, there have emerged two important problems. First of all the special problem involved in shifting the demand for foreign final goods to a demand for intermediate goods was not taken into account. Secondly, the lag between the emergence of a potential market for import substitutes and the effective decision of the private sector to satisfy that market was ignored (Izquierdo, 1964). In addition the divergent goals of the public and private sectors—and of the pressure groups within each sector—added other complex factors that obstructed the rational practice of protectionism.

Furthermore, Izquierdo notes that the government has been concerned by the pressures on the balance of payments resulting from the cumulative increase in costs and prices wrought by the restrictive measures, which put them above those prevailing on world markets. The alternative, he maintains, is to act decisively on private investment. The restrictive measures continue to be a regulating factor, for their indiscriminate use would cut off the supply of necessary goods for industry with subsequent damage to domestic demand (Izquierdo, 1964).

There is agreement on the need to revise protectionist policy. For Izquierdo, the subordination of an import policy to an over–all plan for investment in industry cannot be delayed. This need inevitably arises when instruments of protection operate independently, unrelated to any long–range objectives (Izquierdo, 1964, chap. 4, p. 283). A long–term plan would divert the evaluation of the two types of restrictive measures—permits and tariffs—from the examination of their immediate repercussions. At the same time, a closer coordination between the Secretariat of Industry and Commerce and the Secretariat of Finance is required; these secretariats administer import permits and tariffs. The measures proposed by García Reynoso also show a general feeling that has emerged recently. In a broad sense, they are oriented to the solution of the distortions he analyzes, based on the continued promotion of industry. More specifically, he calls for the revision of the protectionist mechanism and points out that the system of permits should be subject to considerations about the level of prices, the use of domestic inputs, and the effect on the balance of payments (1968).

Some interesting conclusions can be inferred from the works on protectionism. In the first place, it is accepted, without exception, that the protectionist policy has been indispensable for the growth of industry. While its shortcomings are readily admitted, this does not invalidate the fact that it has been the main instrument for the establishment of the industrial apparatus. In other words, the issue was whether or not to create an industry, and this issue was very important for nationalist economic policy. In the second place, in the late 1950s, when the rate of import substitution began to decline, there existed a basic industrial framework, the country entered a period of exchange stability, and pressures on prices were lessened; a new subject of increasing interest arose—industrial efficiency, measured in terms of costs in relation to world prices. Last but not least, there is no talk of eliminating protectionism in order to attain efficiency and confront the domestic producer with world competition. Instead, it is considered a matter of perfecting the method of licensing and selecting imports in accordance with the broader scope of industrial integration, linked to an investment policy.

It is worth noting that no analysis has been made of the role of direct foreign investment in the process of import substitution and, consequently, how foreign firms have benefited from the protectionist trade policy—a benefit that goes against the nationalist interests of the policy's makers. Lastly, the idea that industry has still not matured fully is maintained implicitly. It is noted that Latin American economic integration can provide stimuli on the market side and could be a means of selecting the most efficient firms.

IV. *The Distribution of Income*

The distribution of income shifted in a regressive way during World War II, because of the diminishing relative share of wages and salaries. The possible causes of redistribution, different in their nature and effects, that attracted the attention of Mexican economists above all were agrarian reform and inflation. The interest in the latter quickly surpassed that in the former. In the 1950s—a period of rising prices—numerous and interesting works appeared on this subject, including an essay by I. M. de Navarrete, in 1960, which is the best known work on the distribution of income in Mexico. It is unanimously agreed that there is a very inequitable distribution of income, even more so than formerly. Different observers present different reasons for this, but some points in common can be found. Initially, the following section will present different viewpoints on the inequitable distribution of income. The causal factors will be examined next. Finally, some of the measures proposed to correct the inequitable distribution will be considered.

In her study, Mrs. Navarrete asserts that the distribution of income was more inequitable in 1960 than it had been twenty-five years before. That is to say, the salaried and farming sectors have experienced a decrease in their relative share within national income. This assertion is supported by the conclusions of a mission from the World Bank which established the same point for the period 1939–1950 (Birf y Nacional Financiera, 1953). The same writer also indicates that there is a process of concentration which only benefits capital owners. These opinions are shared by almost all of the economists who have dealt with the subject, such as V. Urquidi, Navarrete, J. Noyola, Flores de la Peña, Carmona, B. Siegel and others.

Because of the inflationary method followed by the government, development policy tends to accentuate the inequitable distribution of income. Such is the opinion of various authors, including Mrs. Navarrete (1955). Moreover, these authors believe that this is inevitable, for in an initial stage of development it is necessary for economic policy to foster the formation of capital. Only at a later stage will the efforts directed at enlarging the domestic market solve the inequitability. As it will be shown later, many feel that Mexico has already reached this stage. In short, the consensus of opinion is that there has been a process of concentration of income in favor of profit earners. This concept, which was dominant during the inflationary period, has not been refuted in the recent years of exchange stability.

Several explanations are offered for this type of income distribution. Actually, the various opinions supplement each other and center on economic and institutional factors. The first group to be presented here is that concerned with the effects of the economic variables involved in the production function.

In the early 1950s, two complementary articles appeared in *Trimestre Económico,* which approach the problem from the standpoint of salaries. One was a joint study by Noyola and D. López Rosado (1951), and the second was by Flores de la Peña and A. Ferrer (1951). The former two pointed out that the level of real salaries had deteriorated, despite the increase in the real per capita income of the country and in productive efficiency in every activity, but that the income derived from this increase had gone to the holders of capital. The determining factor had been a surplus in the supply of labor (excess supply). The other two authors mentioned above note that the owners of capital have experienced additional gains through inflation, since they kept the profits generated by the increased productivity. Urquidi (1961) of the structuralist school, derives the inequitable income distribution from the characteristics and effects of the supply conditions of some factors of production, such as the rigidity of agricultural output, the insufficient rate of industrialization, and the high rate of growth of the supply of labor. Finally, Mrs. Navarrete (1959) also attributes the disparity of incomes to low productivity, which can be explained by the fact that within the occupational structure, there is a predominance of agriculture, handicrafts and small proprietors.

Included among the economic explanations are inflation, the characteristics of the tax system, and the situation in world markets.

According to the consensus, the inflationary process had a redistributive effect in favor of the owners of capital, although no empirical proof is presented to substantiate the statement. What is really being pointed out is that the rise in prices accentuates the inequitable distribution or constitutes an obstacle to corrective measures. No one contends that without inflation there would be an equitable distribution. The characteristics of the tax system are singled out as another causal factor, taking into account that it is regressive. The tax burden is not shared equitably among taxpayers. The predominance of taxes on expenditures and the absence of taxes on investment of limited social value prevent the tax system from acting as a mechanism of redistribution of income. Many experts agree on this, for example, Mrs. Navarrete, Urquidi, Flores de la Peña and others. They all attribute the situation to the fact that the objective of equitable taxation has not been applied properly by the fiscal authorities, which have been mainly concerned with tax collection. There are also those who link the problems of income distribution to the conditions of world markets, like Flores de la Peña who maintains that in export activities the transfer of income from the labor sector to the capital sector is promoted sometimes by currency devaluations; another factor which also has an adverse effect is the price increases in the world markets. (Flores de la Peña & Ferrer, 1951, p. 621). Last of all are the so–called institutional factors. These can be summarized fundamentally by the lack of organization among workers or their limited bargaining power. Moreover, after 1940, wage policy ceased to be as aggressive as it had been formerly, which also helps to explain the low level of real wages and the advantageous position of capital. On this topic the absence of disagreement among Mexican economists is evident. They agree on the skewness of income distribution, and their opinions about the causes of the unequal distribution are complementary. There is one question remaining: how to mitigate the inequality.

Naturally, most of the authors argue for a better distribution of income and advocate action on what have been termed causal factors. In their attempt to achieve this objective, they want a preponderant role for government measures. The intervention of the state for the promotion of economic development, applicable to the productive apparatus, would abolish unemployment. The concomitant increase in productivity, together with monetary stability, would be to the advantage of real wages. Such is the opinion of Navarrete (1955) and others who are concerned about economic growth. The preferred instrument, undoubtedly, is fiscal policy (despite the fact that it has been shown that it is very limited as a mechanism of redistribution in industrialized countries). Mrs. Navarrete considers that because of the disparities generated by inflation, fiscal policy should absorb a substantial part of profits and channel them toward the kind of investment that raises productivity (1955, p. 245). In the early 1950s, a radical reform of the tax system was sought, although now the emphasis is on improving and perfecting the system. In 1951, Flores de la Peña and Ferrer (1951) were clamoring for the elimination of the regressive nature of taxation by means of an increase in direct taxes. Six years later, Urquidi [1956] urged the same thing. The former of these authors also recommended the establishment of additional taxes on exports to absorb the exchange profits of exporters, which increase considerably in the event of a devaluation. As a direct measure to increase the income of wage earners, they proposed a more aggressive wage policy and a greater mobility of the wage scale. Similarly, references are made to the objective of improving the bargaining position of wage earners. These opinions are expressed by Mrs. Navarrete and Flores de la Peña, but they do not propose concrete measures for the removal of the institutional factors that promote an inequitable income distribution.

Most of the proposed measures would either speed up or delay economic growth, depending on the point of view one has of economic development. In other words, whether the stimulus for investment must come from saving, fostered by raising the level of profits, or from demand, fostered by enlarging the market to guarantee the absorption of the goods produced and based on mass consumption. For some, then, the inequitable distribution of income guarantees the process of capital formation. Others, however, closer to Keynesian ideas, believe that the development of Mexico is now limited by the insufficiency of the domestic market. A better distribution of the national income, besides meeting the requirements of social justice, is warranted by the objective of accelerating the process of growth.

The argument for the enlargement of the domestic market rests on the inequitable distribution of income and on the limited demand for consumer goods industries, some of which do not utilize their full capacity. The first of these reasons has to do with distribution, while the second is related to production. The evaluation inferred from this situation will depend on the predominant goals of economic policy. Proof of this can be found in recent statements of the Minister of Finance, A. Ortiz Mena, who asserts that fiscal policy must be defined in terms of the formation of the domestic market, without which economic development cannot acquire sustained momentum. A healthy and fair development, he maintains, requires that the inequalities of income be corrected (1966, p. 9).

Among Mexican economists, this standpoint is not new. Combined with many elements of Keynesian policy, it already had a following during the inflationary period, which was characterized as stemming from an excess demand. For example, *Revista de Economía,* representative of current economic thought, in its editorial of May, 1954, pointed out that the crucial problems of domestic economic development have as their basic reason the constant contraction of the domestic market because of the smaller relative incomes of the groups with a high propensity to consume. If that is so, the effects of the peso devaluation (1954) and future policies must be evaluated and conceived on the basis of this conditioning factor. If the premises are valid, the future policy of the Government must reenforce as much as possible the purchasing power of the masses of the national population, not only for reasons of social justice and equality, but as a central foundation for a theory of economic growth. Urquidi agrees with this idea and also mentions the insufficiency of the domestic market as a result of the structure of the distribution of income. The concentration of income, due to inflation and structural elements, results in a structure of total demand which does not seem congruent with the possibilities of full employment of the capital of the country (Urquidi, 1956, p. 435). There, then, is one reason to enlarge the domestic market. Summing up, this idea has been shared by a great number of economists. It has also been the basis for the explanation of inflation expounded by the structuralists, as will be shown later. There is a divergence of opinion on the same subject. For instance, R. Hoyo D'Addona (1968) states that in Mexico there is still an excessive potential demand in terms of total supply, and that therefore, the fiscal system should stress the capitalization of the country.

Another phenomenon that has been observed within the structure of demand has been the change experienced as a result of increasing urbanization. According to Fernandez Hurtado (1960), the exodus from the rural areas to the urban centers was due to the agrarian reform. Yet, whether it was prompted by the armed struggle or the distribution of land, it had no economic justification such as would have been provided by an increase in the productivity of land and a subsequent excess of agricultural output relative to rural consumption. The increase in the urban population has marked a new direction in the structure of the demand for the goods of domestic industry and has generated an extraordinary boom in the urban construction industry, a net consumer of national goods (Fernandez Hurtado, 1960, p. 202).

V. *Stabilization Policies*

Monetary Aspects

Like other variables of the system, the monetary behavior of Mexico since World War II can be divided into two distinct phases. The first, until 1957, was marked by inflation; the second, from 1957 to date, has been characterized by stability. An inflationary process began in the Mexican economy in 1935, during which time money grew proportionately to income. From that time until 1956, the average annual increase in prices was 11 percent, although with wide fluctuations, especially during the war and the periods immediately following devaluations. However, in the period between 1956 and 1968, price increases have decreased asymptotically and have become moderate. Similarly, changes in the money supply are now more than proportionally related to the growth of money income.

In the beginning, the inflation had mixed characteristics of cost increases and increases in demand. Later on, the latter was the predominant characteristic, for the public sector obtained funds by inflationary means and promoted inflation, as its expenditures constituted an excess demand in relation to productive capacity.

Foreign trade was another factor in the process. Exports of agricultural commodities—the only significant exports—grew very rapidly due to price increases and, especially after the agrarian reform and the agricultural development works that followed, also by production increments generating thereby rises in the money supply. Nevertheless, in the period of stability which began in 1957, and was associated with smaller increases of agricultural output, there was a decrease in the rate of growth of these exports. The income-elasticity of the demand for imports also fell from one period to the other, but the policy of protectionism changed the structure of imports so that those of capital goods and intermediate goods became predominant.

The devaluation of 1938 was related to the expropriation of oil; that of 1948–1949 was considered to be part of the worldwide parity changes that took place in the postwar years. On the other hand, the devaluation of 1954 was an isolated event which had a drastic effect on national expectations. Once more it was promoted by public spending, executed by a new administration which, using Keynesian criteria, tried to counteract the decrease in business activity (attributed to the recession in the American economy in 1953 after the Korean War) by means of deficit spending financed by the central bank. The devaluation, which was described as a preventive measure, established the exchange rate at its present value of 12.50 pesos to the dollar. Conditions had changed as compared with the preceding devaluations. The system had become more flexible in terms of output especially in the agricultural sector. This, together with the more extensive use of foreign loans for public investment after the devaluation, meant that the central bank no longer had to finance the investment of the public sector. Thus the depreciation of 1954 was a transitory inflationary influence, since the economic variables that affected the level of prices followed a trend toward equilibrium and determined conditions of growth with price stability. Later on, beginning in 1958, saving in financial institutions tended to grow more rapidly, as price expectations changed, so that it was possible to finance the new government deficits with private savings, placing government securities in the banking system to supplement foreign loans. In this process, the greater use of reserve requirements was very important. This instrument was also applied to non–monetary intermediaries, which were the institutions that registered the fastest growth. It also made it easier for the Banco de Mexico to attract genuine savings to finance public investment. To recapitulate, the former monetary expansions were replaced by the use of domestic and foreign savings. The earlier public investment contributed to higher output, and there was also a change in the structure of consumption and an increase in the propensity to save. In effect, the initial elimination of the government deficit, and the different way of financing it employed later, abated the inflationary demand pressures stemming from public finance.

The Inflationary Period

The inflationary process has produced a diversity of opinions and controversies about monetary policy. In the first part of this section, attention will be given to the interpretations of the monetary phenomena that occurred during the inflationary period, followed by the different explanations of the causes of the inflationary situation itself. In order to facilitate the presentation, events will be grouped together in two chronological periods: those which center on the devaluation of 1948–1949 and those which took place around the time of the devaluation of 1954. The dividing line has been established as the Korean War, in 1950, for it put an end to the effects of the first devaluation and began to generate the causes of the second.

For a majority of the experts, the inflationary process immediately following World War II was the result of two factors. The first of these was the strong drive toward public investment with deficit financing. The second factor was the growing deficit in the balance of payments, attributed to the excess of imports that corresponded to the deferred demand of the war years.[5]

---

[5] For some authors, however, the second factor was of an external nature—the end of World War II.

Were there other conditions that led to the devaluation of 1948–1949? A classic liberal viewpoint was expounded by a banker, A. de Iturbide (quoted by E. Padilla, 1948, p. 397), who pointed out that State intervention was a cause of the disequilibrium for it brought about reforms that resulted only in price increases and decreased agricultural output. Among other factors were: the agrarian reform and financing *ejidatarios* and small proprietors; the aid given to production cooperatives; and the oil expropriation and the nationalization of the railroads. Another viewpoint was based on the application of the purchasing power parity theory to the relative levels of prices in the United States and Mexico. Saenz is representative of this current of thought, and he adds to this analysis the influence of the stock of money. Refuting an argument of the Economic Commission for Latin America (ECLA), he stated that the peso had been overvalued since 1941, and that this fostered imports and the balance of payments disequilibrium. If Mexico did not devalue before (1948) it was for reasons derived from the war: a) the control of production in other countries; b) the amount of capital which found a haven in Mexico and meant an abundance of dollars despite the exchange disequilibrium, and c) the remittances of dollars by Mexican farm labor in the United States, the sum of which was as important as exports (E. Padilla, 1948, p. 405).

The devaluation did not occur until 1948 because of the Government's fear of the consequences. It did not want to devalue before industrialists could purchase abroad at low prices, which they had not been able to do during the war. Another reason was the case of obtaining foreign loans, and still another was the hope that a point of parity would be reached autonomously as a result of the inflation in the United States (Sáenz, quoted in E. Padilla, 1948). According to Gómez, Director of the Banco de México (1964), the overvaluation of the peso was brought about because prices rose more rapidly than in the United States between 1945 and 1948, and when price controls were lifted in that country, it was impossible to correct the disequilibrium in the balance of payments by reducing domestic demand. He added that the greatest effort was made to maintain the exchange rate by using reserves and foreign credit. But the offers of further credits required a balanced budget, which was impossible in the short run. For this reason, there finally was a devaluation (Gomez, 1964, pp. 778–779). In an attempt to place the blame for the devaluation, Sáenz holds all of the economic groups responsible. The blame was shared by the government with its equivocal monetary policy which made production costlier; the private sectors because of their policies of hoarding and of speculation, and especially the industrialists, for not improving technology and relying on the support of excessive protectionism; labor for demanding high salaries; and private banking, because of the voracity of some bankers. In short, the situation brought about a rigid cost structure, which could not be reduced, far above that which should have prevailed with an exchange rate of 4.85 (E. Padilla, 1948, pp. 406–407). Finally, he adds, monetary policy between 1945 and 1948 was equivocal, in allowing the dwindling of reserves in the interest of maintaining the exchange rate and because the new exchange rate only left the peso with a very slight undervaluation, which was quickly neutralized. What Saenz therefore recommended was a more austere monetary and credit policy, to be pursued within more restricted margins, since the currency was chronically overvalued with respect to the rate of exchange. This made it necessary to continuously limit the quantity of money in order to avoid serious repercussions on the reserves of foreign exchange (Saenz, 1957). It would also be necessary to revise protectionist policy, for excessive protection distorts the struture of domestic costs and thereby fosters overvaluation and price instability. Therefore, this policy, including tariffs and subsidies, permits and quotas, should be administered much more carefully than monetary and credit policy (Saenz, 1957, p. 538).

Another current of economic opinion is to be found in the opposition from the left. Often they introduce external factors as causes for the disequilibria and are strong partisans of exchange controls. According to E. Lobato Lopez (1968), the orientation of Mexican economic policy has always been predominantly Keynesian and monetary policy was actually a policy of forced saving, which concentrated collective saving into a few hands, in order to encourage production at the expense of consumption. N. Bassols (quoted in E. Padilla, 1948) focused on political and external factors. He maintained that if the inflation was not constrained at the right time, it was for political reasons and because all hope converged on foreign aid–loans or foreign investment. He felt that this policy should be abolished and than attention should be paid to the country and her internal solutions. The same line of reasoning is followed by R. Torres Gaitán who emphasizes external factors, and more specifically the attitude of the United States. He notes that American purchases during the war brought about an increase in the money supply which led to inflation. Later on, when the United States began to sell to Mexico, problems emerged in the form of balance of payments deficits. Therefore, Mexico should have revaluated her currency during the war, and once the war was over, devaluated it (quoted in E. Padilla, 1948, p. 402). Noyola (1948) made the observation that the rise in prices stemmed from the demand side as well as the supply side. On the demand side, the two outstanding factors were the growth of the money supply and deferred demand. The pressures generated from the supply side pertained to the increases in the prices of imported goods, due to their scarcity because of restrictions on imports; the rise of prices in the United States, and the effects of the devaluation of 1948.

These economists view the devaluation in a different light. According to Bassols, the monetary reserve was not used in the best possible way, that is, to purchase production goods, instead of which much of it was squandered (quoted in E. Padilla, 1948). Torres Gaitán distinguishes between external and internal causes of the devaluation: the external factors are based on the decadence of capitalism and the policies of the United States and the IMF. Those of an internal nature include the balance of payments disequilibrium, the price differentials between the United States and Mexico, the flight of capital, the deferred demand for imports, and the complete failure of import controls. Of all of these, the disequilibrium of the balance of payments was the most important. In any case, he concludes, the devaluation was extemporaneous (in E. Padilla, 1948). Finally, A. Noriega Herrera (1955), indicates that the short–term results of the devaluation were positive, but he casts a doubt on the timeliness of the measure, asking whether it would not have been preferable to establish exchange controls or whether the monetary authorities carried out the devaluation, believing in the "optimistic promises of the International Monetary Fund." The negative results readily appeared, since the price of imports increased in domestic currency, commodities became more expensive for the domestic consumer because of the preference to sell export goods abroad, and inflationary pressures emerged in all economic activities (Noriega, 1955, p. 159).

The solutions proposed by Torres Gaitán can be considered representative of this current of thought. He maintains that if the country's exports could be increased, and if imports could be reduced, not only the balance of payments would be righted, but also domestic activity would be stimulated as a consequence of more benefits stemming from the increase in exports and the decrease of goods that can be done without (in E. Padilla, 1948, p. 401). He subsequently proposed price controls, exchange controls, and tailoring foreign trade policy to agricultural and industrial policies. Noyola (1948) was more inclined to find the primordial causes of the price increases in the real factors of the economy. He stated that a policy of price controls would be tenable only if there were controls on the volume of commodities, and not the "police" measures that had been used until now. Moreover, it would be convenient to apply, in a supplementary policy of moving wage scales linked to the cost–of–living index. Nevertheless, success would depend on monetary and fiscal policy and an uncontrollable factor–American inflation (1948, pp. 9, 10).

Until now, the opinions presented have concerned the events prior to the Korean conflict. Opinions on the events that led to the devaluation of 1954 will be examined next, and then the end of the inflationary process and the stability attained between 1957 and 1958 will be reviewed.

Gómez, an exponent of the official view, interprets the devaluation of 1954 in the following way. The excessive credit generated by the banking system to finance the public sector deficit, as a result of following Keynesian–like advice to offset the effects of the 1953 American recession, drained the foreign exchange holdings of the central bank. But on this occasion the policy pursued was different from that of 1948—the peso was devalued before the monetary reserves were depleted and before resorting to greater indebtedness to the IMF and the U. S. Treasury (1964, p. 780). On the other hand, W. Sedwitz (1957) stresses phenomena in the external sector that preceded and provoked the devaluation. What precipitated the devaluation of 1954, he claims, was not the absolute level of the foreign exchange reserves, but the rate at which they were dwindling (1957, p. 12). The deterioration was due to a) the decline in the volume and the price of exports, as well as to b) the slowdown in business activity in the United States. These were the two most important factors. c) Moreover, the deterioration was accompanied by an increase in imports as a result of the credit and fiscal expansion which constituted an attempt to offset the effects of the American recession; and d) the situation was worsened by the outflow of short–term capital, made possible by internal liquidity and the fear of devaluation or exchange controls. The outflow of capital in 1954 was promoted by the excessive liquidity of the monetary system and by the ease of selling government securities without capital losses and of placing profits into dollar accounts in Mexican or American banks (1957, p. 23). This is why a restrictive–depressive policy was difficult to maintain and perhaps the monetary authorities themselves wanted to count on wider margins to carry out the anti–cyclical and development policies. According to Sedwitz, it would be unfair to infer that the devaluation was equivocal, since the positive phenomena that followed had still not taken place and because the pressures on the dwindling reserves of foreign exchange and domestic depression were very strong.

The leftist current also made new contributions. It is possible that it was strongly influenced by the publications of the ECLA during that period, which had not existed at the time of the previous devaluation. The reference here is to concepts that later gained wide acceptance, such as the theory of the deterioration of the terms of trade of Latin America and the dualistic approach of R. Prebisch (ECLA, 1950), who divides countries into two large groups—the cyclical centers and the peripheral nations. These economists continued to stress the external sector, the idea of a cyclical center providing an explanation for the origin of all the cyclical disturbances in the peripheral countries. The views expressed by Noriega (1955) are an example of this approach. To him, the Korean War was the reason for the disruption of the relative stability that had been attained, for it brought about a series of phenomena that altered the entire Mexican economic system. Among the most important changes were the inflow of foreign capital, the rise in the world prices of raw materials, and considerable increases in receipts of some invisible items of the balance of payments, that cooperated to trigger domestic inflationary pressures. Among internal factors he stresses the lack of skills of those in charge of economic policy and the absence of a measure to limit the remittance of profits by subsidiaries of foreign companies to their countries of origin. The measures chosen to abate the inflation were not successful because of the new trends in the countries that are cyclical centers, such as rearmament, the speculative boom, the contraction in the raw materials market, and the increased tax burden (Noriega, 1955). Concerning the balance of payments disequilibrium, he pointed out that it was a result of the worsening of the terms of trade, since export prices fell 14 percent while the prices of imports rose 0.8 percent (1955, p. 163). The solutions offered do not differ from those that had been proposed for the devaluation of 1948. Thus, this writer stated that devaluations have proved incapable of correcting the periodic disequilibria of the balance of payments, much less of offsetting them. The solution lies in multiple exchange controls, which he considers advantageous because they limit the profits of exporters and adjust imports to the available foreign exchange. They also protect the cost of living, permit the increase of the revenues of the State, and encourage tourism and new exports (1955, p. 168).

Taking a position contrary to official policy (and within the field of the structuralist current) Lobato (1968) stated that inflation had two fundamental effects—price increases and a reduction in the purchasing power of the majority of the population. Economic policy is not effective for it cannot reduce prices or even curb price increases, nor can it augment the standard of living of the population. Since taking action on prices is difficult, the solution would have to be the redistribution of income, by means of a decisive policy of raising wages and salaries. Technically, he says, this policy is valid, for it would enlarge the domestic market and foster development.

The inflation gave rise to different viewpoints on its causes as well as on appropriate solutions. Opinions were very definite, frequently based on the writings of a favorite author. There was one current of opinion, the structuralist, opposed to public policy, which formulated a novel explanation of the inflation and which when its proponents were joined by other Latin American economists came to represent a new "school" of thought. Others remained within the bounds of an orthodox interpretation, even though they disagreed on the importance of the determining variables.

The Structuralist Position: The most outstanding spokesman of structuralist thought was Noyola, who was the first to systematize all the pertinent tools of analysis in 1956. The current was developed in conjunction with the staff of economists of the ECLA, in Santiago de Chile. The work of M. Kalecki at the United Nations Secretariat was influential upon the economists that rallied to structuralism. The controversy between ECLA and the IMF about inflation in Latin America influenced the way of thinking of many.

Noyola readily admits the influence of two European economists, Kalecki and H. Aujac. The former underlines the importance of the rigidity of supply and the extent of monopoly power in the economic system. In contrast, Aujac focuses on the behavior of the various social classes and their bargaining power. Noyola summarizes his position briefly: inflation is not a monetary phenomenon; it is the result of disequilibria of a very real nature that are expressed in the form of increases in the general level of prices (1956, p. 604). He introduces a series of factors to examine inflation in Latin America and divides them into three groups: those of a structural nature (population and productivity); those of a dynamic nature (different rates of growth in the diverse sectors of production); and those of an institutional character (the behavior of the public and the private sectors).

Noyola points out that among the factors which produce inflation there are two distinct kinds. One of them is made up of what he calls the "basic inflationary pressures," on the magnitude of which depends the degree of inflation. These pressures are due to disequilibria of growth, which are almost always to be found in the slow growth of two sectors, foreign trade and agriculture. The second group consists of "mechanisms of propagation" which have secondary or spread effects, for they only heighten the inflationary process through the performance of the economic system itself. The mechanisms of propagation can be of three kinds—the fiscal mechanism, the credit mechanism, and the mechanism for the readjustment of prices and incomes. It can be said, then, that the inflationary process is born from the emergence of the basic inflationary pressures that result from the disequilibria arising from the efforts of growth in the economic system. Once the inflationary conditions are established, inflation is further stimulated by the mechanisms of propagation. Some conclusions inferred from the model are applicable to the Mexican reality. Since 1935, the basic inflationary pressures in Mexico have stemmed from the inability of exports to grow at the same speed as the domestic economy. This has led to disequilibria in the balance of payments and to the subsequent devaluations (Noyola, 1956, p. 611). Yet, the increasing diversification of exports has abated the force of these pressures. With regard to agriculture, Noyola inferred something else: that there was almost no inflationary pressure stemming from the rigidity of agricultural supply, that is, the supply of foodstuffs (1956, p. 612). This was the result of the agrarian reform and the agricultural development policy. However, in the short run, agriculture had played a role in that sense, for example, when the crops had been poor (Noyola, 1956).

The distribution of income and the behavior of social groups are manifest in the mechanisms of propagation. Noyola remarks that in the case of Mexico these factors have been weak, and that the weakness has been even more pronounced among the lower-income groups. Taxation had tended to become more regressive because of the increases in prices, and the role of current expenditures as a redistributor of national income has been insufficient. The credit mechanism had been the most passive in the propagation of inflation, which only went to prove the well-known axiom that monetary policy is only effective when it abates business activity, increases unemployment or curbs development.

The mechanism for the readjustment of prices and incomes was the most important of the three in propagating inflation, because of the high degree of monopoly that exists in the Mexican economy. With the exception of rental housing properties, in almost all other prices of goods and services, entrepreneurs were able to transfer the incidence of inflationary pressures, although salaried groups could not do the same (Noyola, 1956, p. 614). This proved the weakness in the adjustment of salaries, which stems from structural and institutional reasons. The structural reason was to be found in the existence of a surplus army of agricultural labor with a very low productivity, which tended to depress the level of real wages and to debilitate labor organizations (Noyola, 1956, p. 615). The institutional reason lay in the relationship between organized labor and the public sector, that is, government paternalism that weakened the labor movement. To recapitulate, Noyola identified two factors as the principal causes of inflation in Mexico—the foreign trade disequilibrium due to insufficient exports as the basic inflationary pressure, and the degree of monopoly as the mechanism of propagation. His conclusion was that if a choice must be made between inflation and economic stagnation or unemployment, inflation is preferable. The negative side of inflation is not the increase in prices, but its effects on the distribution of income and the distortions that it wreaks in the productive structure and the structure of demand. It may be impossible to restrain inflationary pressures, but they can be abated by means of a very progressive fiscal policy, as well as price contols and wage adjustments (Noyola, 1956, p. 616).

Another author, Flores de la Peña (1953), introduced a few other elements in his analysis of the Mexican inflation. He differed from Noyola in that he considered the inelasticity of agricultural production as one of the determining factors of instability. He also identified, as an inflationary pressure, the increase in public investment, which he considered to have had the opposite effect than Noyola propounded, i.e., it fostered the expansion of output. Flores de la Peña differed again on the issue of the role of exports, for he felt that the growth of exports constituted another source of inflationary pressures, because it brought about the boom in foreign investment, a greater sensitivity to international economic contractions, and an inadequate specialization in basic commodities. According to Flores de la Peña, the growth of exports brought about an increase in the quantity of money, and therefore, had an inflationary effect. Noyola, however, considered that it was the insufficiency of exports vis-a-vis the increase in imports that created the balance of payments disequilibria that led to the devaluations and the changes in the level of prices. Within Noyola's reasoning, there is evidence of the theory of the deterioration of the exchange rate by external reasons, such as the decline in the prices of exports. Flores de la Peña proved that the effect of the balance of payments on inflationary pressures had not been significant. He maintained that the influence of the increase in prices abroad or the devaluations on the general price level had been overstated. He felt that it ought to be admitted that the inflationary pressures had been caused by gross investment, or, especially, by its composition (1953, p. 470). It is precisely on investment that he based his explanation of the inflationary process. Public investment created inflationary pressures because of the lag between its monetary effect and its real effect. Private investment creates these pressures when it is "poorly oriented," that is, when it is directed toward superfluous goods or for foreign markets. In general, "the expansive capacity of investment will be greater as the proportion of income generated passing to the lower-income groups is greater" (1953, p. 467). Besides, the origin of inflationary pressures lies in the low productivity of private and public investment. The inelasticity of supply plays a role in the mechanics of inflation, according to Flores de la Peña. The inelasticity of supply is due to the inelasticity of domestic output and the scarcity of foreign exchange. The reasons for the former have already been examined in the section on agriculture. The scarcity of foreign exchange is a phenomenon that is not uncommon in a period when there are substantial imports of capital goods, which prevents covering the deficiency of output with commodity imports. Consequently, the inflationary process due to economic development begins when an increase in the income generated by gross investment raises effective demand to the extent that it brings about a scarcity of factors of production. In an underdeveloped country, this point is easily reached because of the inelasticity of supply (1953, p. 477). Thus, inflation can break out before the economy reaches full employment, due to the fact that any increase of income gives rise to a more than proportional increase in the demand for foodstuffs and manufactured goods for mass consumption, and rarely does the growth of investment mean a significant expansion of the output of these goods (1953, p. 462).

Until now only the differences of opinion of the two authors have been discussed, but they share some opinions. Their similarity lies in that they both examine inflation in its real form. They reject the monetary approach and monetary policies and they imply that inflation is inevitable in the process of development. Both have been influenced by Kalecki, and both share a concern about the effects of inflation on the distribution of income. In other words, both authors study inflation by comparing total effective demand with total supply, stressing the real aspect. They consider that monetary policies have not been effective because they could not prevent the increase in the money supply. According to Flores de la Peña, monetarist policy seeks a balanced budget and prefers monetary and credit policies, to foster economic development. But it must be understood that the quantity of money is merely a result and that by itself it does not determine anything. The solution is simple: reduce demand or increase supply. Thus, monetary policy only affects demand and as a means of reaching equilibrium is an absurdity in a country that has a very limited domestic demand (1953, pp. 472–473). These ideas supplement those of Noyola, who states that monetary policy is effective only when it strangles economic development (1956, p. 614). The solutions given by Flores de la Peña are directed basically to acting upon investment. Thus, the intervention of the State acquires a decisive function, for it must establish priorities for investments, trying to maintain the equilibrium between effective demand and total supply. Moreover, for the reestablishment of equilibrium, he advises finishing the agrarian reform in order to improve agricultural production and eliminating the insufficiency in the supply of foodstuffs. He suggests that raising wages and salaries is useless because prices would increase at a faster rate, and that the only way of maintaining the purchasing power of the income of the urban population is by increasing the supply of foodstuffs and manufactured goods for mass consumption. If the process of the concentration of income is allowed to go too far, a point would be reached where any reduction of the real wage would be accompanied by a more than proportional decrease in the effective demand for manufactured goods. He asserts that this has happened in Mexico in basic industries, like the textiles and shoe industries. Flores de la Peña does not see any advantage in price controls, because decrees do not increase output, and it is precisely upon output that economic policy must have an effect (1953, pp. 479–480).

Although it is difficult to really place him within any current of thought, Torres Gaitán (1962), presents a viewpoint considerably different from those of the other two authors. Acknowledging that the external disequilibrium is the result of a domestic demand which exceeds domestic supply, he feels that the solution lies in increasing the latter as well as raising national productivity. The insufficiency of domestic supply helps him to explain both the deterioration of the exchange rate and the basic inflationary pressures, in the same way as the structuralist approach. Torres Gaitán adds that the adjustment has been made by means of different instruments. Before World War II, the main instrument was beggar–my–neighbor policies. More recently, the most important instruments have been controls on foreign trade and the use of foreign loans. He points out specifically, moreover, that the permanent disequilibrium of the exchange rate is due to exports of basic commodities with a negative income elasticity of demand made more acute by the encroachment of synthetic substitutes, and by the excess supply of primary commodities that depresses their prices (1962).

In its analysis of external equilibrium in Latin America and regional foreign trade problems, the ECLA studied the case of Mexico, with the collaboration of Noyola, C. Furtado, O. Sunkel and Urquidi. In synthesis, they asserted that the devaluations had occurred not because the exchange rate was overvalued, but for structural reasons. Structural reasons predominate as a result of economic development itself, that is, the incongruence between the pace of increase of total supply and total demand. The disequilibrium is even more acute when development is accompanied by a concentration of income, due to the high income–elasticity of imports of durable goods (United Nations Economic and Social Council, 1957).

In practice, they present the import function as dependent on the rate of growth of income, which tends to produce foreign disequilibrium when there is an acceleration of income growth; disequilibrium is accentuated by the deterioration in the terms of trade (United Nations . . . , 1957). Devaluations do not alter this combination of causal factors, because it is a structural phenomenon of the economy, related to the inelasticity of supply (especially that of agriculture), which plays a role in making the response to increases in external demand quite slow. Nevertheless, the higher foreign exchange income is reflected in higher domestic spending, price increases and increases in imports, until the value of exports falls because of a change in the terms of trade, and the current account deficit is heightened. Thus, if devaluations were effective in the past, for example, in stimulating exports of certain goods, such as cotton, or in fostering the substitution of imports of other goods, later on the range of these possibilities became much narrower (United Nations, 1957).

The next part of the analysis was the object of controversy. In order to prove that the external disequilibrium had structural causes, it was deduced, from a series of price indexes based on the purchasing power parity theory, that the exchange rate had been undervalued since 1937. In other words, the external disequilibrium could not be attributed to an overvalued exchange rate.

The Orthodox Position: Sáenz (1957) disagreed with the thesis of ECLA. He considered that the overvaluation of the rate of exchange during that period had been the central problem. Besides pointing out that ECLA should have defined what was meant by an equilibrium exchange rate, he noted that between 1937 and 1955 the balance of payments had undergone periods of equilibrium and dynamic stability, which did not appear in the tables elaborated by that organization. On this basis, Sáenz maintained that it was wrong to speak of a period of overvaluation or undervaluation between 1937 and 1955. Moreover, the ECLA thesis has some technical shortcomings. Within the theory of purchasing power parity, it is misleading to use a very long period of time because of the variables that are omitted from the estimation, which with time may become more important than those that have been included (1957). Similarly, he asserted, ECLA used a base year which was not representative. If the periods had been divided into more logical stages, the conclusions would be different. Besides, greater refinements should have been used; for example, sectorial and special price indices, changes in the degree of tariff protection or variations in the elasticities of exports and imports (1957). In conclusion, the data, presented in a more orderly fashion, would indicate that the peso was overvalued in the period under consideration and for that reason it has been more predisposed to shocks or disturbances. For Sáenz, the increase in prices did not stem from the absence of a balanced budget, except to a small extent, for the expenditures had been held at prudent levels. The real causal factors were the generation of income from development itself, the expansion of bank and extra bank credit, and the general increase in costs due to commercial policy. The first and the last of these were inevitable (Sáenz, 1958).

One group of economists considered that external factors are fundamental in explaining monetary disequilibria. Some of them shared the thesis of ECLA, but without accepting the production rigidities and while considering the influence of monetary variables. The argument was that public expenditures, or a surplus on the current account of the balance of payments, produces an increase in aggregate demand, and induces increases in investment and income, and that imports rise in an even greater proportion than income. Since the growth of exports is slow—and they might even decline because of the deterioration of the terms of trade—a deficit on current account appears. Change in relative prices—domestic and foreign—or a monetary policy oriented toward sustaining output growth, accentuates the balance of payments deficit and it becomes necessary to devalue to restore external equilibrium. This approach utilizes the type of perverse elasticity of demand for imports that was so fashionable among economists after the end of World War II. (Siegel, 1960, and J. S. de Beers, 1953, estimated an income–elasticity of imports of close to 2, from which they inferred that it was necessary to export more, use more foreign loans or devalue periodically. This viewpoint was shared by Navarrete.)

In his analysis of the Mexican inflation, Siegel used orthodox instruments to determine what kind of inflation there had been. Although he is not a Mexican economist, his work, like that of de Beers, influenced the discussion of Mexican economists. This author studied the goods market and the demand for money in order to locate the source of inflationary pressures, both domestic and external. He came to the conclusion that inflation in Mexico was not a cost inflation, and that the Mexican experience stands as proof that industrialization is not necessarily accompanied by a lag in agricultural output, as some structuralists maintain, for agricultural production grew more rapidly than industrial output and population (Siegel, 1960). It can be said that the periods with the largest price increases coincided with the years of the devaluations. In turn, it would seem that the only real pressure on the cost side related to external factors, apart from devaluations, came through raw materials and production goods and was associated with the strong rise of import prices after the war. Pressure on the wage side was insignificant, due to the weak bargaining position of unions. Hence Siegel concluded that the Mexican inflation was not a cost inflation, and that if it was so in any way, it was because of the increase in import prices and not because of differences in the rates of growth of the various sectors (1960, p. 81). There is no doubt that price increases were the result of the effects of demand. The causes of variations in the money supply, Siegel said, indicated the origin of inflationary pressure (1960, p. 82) and can be identified as internal and external. In general terms, it seems that between 1941 and 1948, internal factors were mainly responsible.

The internal factors most important in generating inflationary pressures were the government deficits and the loans granted by state–owned banks. The deficits were a direct result of the inability of the government to finance its investment programs by means of taxation or non–bank loans (Siegel, 1960, p. 88). It is also true that inflation was tied to development because of the maintenance of a very high rate of investment. Although they were not in themselves inflationary forces, there were institutional and cultural factors that accelerated the inflationary process, for example, the level of technology or the structure of consumption. Furthermore, the weakness of the instruments of monetary regulation was evident, the reason for this lying in the limited development of the money markets, in a banking system that was not very independent, and in inflationary fiscal policies. Siegal adds that monetary regulation is not effective when the government cannot practice budgetary discipline while instituting a development program. In general, the monetary authorities have been more tolerant of inflation than of deflation.

However, provisionally, and at times decisively, the effect of external factors was felt intensely. Such was the case in the periods 1942–1945, 1950–1951 and 1954–1955.

In the periods of devaluation, the pressure coming from external factors was more important. The devaluations of 1948–1949 and 1954 were preceded by declines in international reserves that resulted from excessive imports (Siegel, 1960, p. 93). Siegel is also of the opinion that between 1951 and 1953 the deterioration of the real terms of trade brought on new problems. In short, the external sector has played a very complex role. On the one hand, it has brought about a growth of income through exports. On the other hand, through imports, it has led to balance of payments deficits.

Regarding the solution of the problem of external disequilibrium, the ECLA group believed that it was highly improbable that new devaluations would have positive effects to counteract the contraction of demand or income. An acceptable policy would be the reorientation of productive resources based on the development of demand and on the capacity to import. This could permit gradual structural changes in supply, directing investment and fostering the required substitution of imports (United Nations, 1957). Devaluation, a measure that has been used to counteract external disequilibrium, is a dangerous instrument for the correction of the balance of payments in the short run. ECLA was also of the opinion that devaluation could not correct the fundamental causes of the disequilibrium. Its immediate and long–run effects could accentuate the inflationary process. Other factors, such as the high rate of domestic spending, the tendency toward budgetary deficit, and the weakness of the instruments of monetary control in the face of international disturbances can cancel out the initial advantages of devaluation.

In general, it can be said that inflation in Mexico did not become a hyperinflation, for the characteristic mechanisms acting to increase velocity did not appear. Nor was there a price–wage spiral of any significant proportion. It was, perhaps, conditions in the labor market itself that prevented hyperinflation.

The Period of Stability

After 1955, there was a decline in the number of works on monetary policy. Those published toward the end of the decade, when stability was well entrenched, still referred to the recent inflationary problem. In effect, as the influence of inflation waned, so did the number and intensity of the writings of economists about monetary stability. This waning continued through the 1960s, and the focus of economic discussion moved to other areas.

Sedwitz (1957) attributes the recovery that began after the devaluation, in the second half of 1954, fundamentally to the expansion of exports, but argues that, contrary to what might have been expected, the devaluation had a very small effect on this turn of events. What elements, then, came into play? The same author points out two. One, external in nature, was the recovery of business activity in the United States, which even resulted in an increase in dollar prices in the world market. The other, of a domestic character, was the extraordinary crops resulting from optimal climatic conditions, especially those of cotton and coffee. The phenomenon can be explained by the fact that there is a high price–elasticity of demand for exports. However, exports are more sensitive to income changes abroad than to prices, for they consist mainly of raw materials. The devaluation had little effect on exports because it did not change their dollar price, or the volume of exports (1957, pp. 14–16). The influence of the devaluation in reducing imports was not important, owing to the low (less than unitary) price–elasticity of the domestic demand and the high elasticity of external supply. There was a substitution effect with regard to durable consumer goods, the demand for which was channeled to national industry. Imports of raw materials increased because of the stimulus provided by the devaluation to local industry. Actually, the change in the composition of imports was due above all to the increase in agricultural output, because of the reduction in imports of foodstuffs, as well as to the greater restrictions on imports (Sedwitz, 1957). This analysis is diametrically opposed to the hypothesis of the inelasticity of agricultural supply of Flores de la Peña.

Some experts arrived at the conclusion that the devaluation did not play a decisive role in the recovery of the economy after 1954. The basic reasons for the recovery were the good crops and the boom of the United States economy. Secondly, there was an increase in tourism, on which the devaluation of 1954 had a positive effect.

Many authors abstained from passing judgment on the prospects for stability, remaining on the level of offering solutions and confining themselves to the short run.

Others foresaw the continuation of the inflation, because the basic causes of the problem were not being checked. However, Sedwitz, in 1957, also noted signs that would guarantee future monetary stability. His reasoning was that economic policy has three fundamental functions related to monetary stability: 1) to prevent the outflow of capital; 2) to maintain the stability of the exchange rate, and 3) to make anti-inflationary policy effective by means of a control on the volume of money and prices. In terms of each of these functions, there was a favorable outlook for the role of economic policy, both from the domestic and the foreign points of view (1957).

According to Gómez, the period of adjustment of the money supply and prices lasted until 1960. In the years of world prosperity between 1955 and 1957, the Banco de Mexico recovered its reserves, capital returned to the country, and there was an increase in foreign and domestic private investment, because of the confidence inspired by free convertibility. With the growth of domestic saving, fostered by stability, a big step was taken in the financing of development. Similarly, public investment in social overhead capital was supported by all of the concurrent events (1964). Transitory difficulties could be solved satisfactorily by resorting to the monetary reserve as an offsetting factor. More recently the increase in bank financing of the private sector has reinforced the downward trend of the rate of interest. Interest rates had risen drastically in real terms when price increases ceased, even though there was not much change in the nominal rates. Gómez adds that in the struggle against inflation, the objective was never stability per se (1964, p. 778). In a developing country, inflationary pressures are harder to overcome because there is a smaller capacity for voluntary saving and for the payment of taxes, at the same time as financing requirements are larger (1964, p. 780). Later on, Gómez mentions an institutional factor influencing the performance of Mexican economic policy. For the monetary authorities, the most difficult pressures to resist are those coming from the public sector. Government officials want to advance as much as possible while they are in power, at the expense of other sectors of the administration (1964, p. 78). This is how budget deficits and inflationary pressures are created. Moreover, state-owned enterprises might be another source of pressure, because of negligence in matters of efficiency and productivity. The pressures caused by the private sector originate mainly in the banking system, because of the banks' interest in obtaining reduced reserve requirements as well as unlimited rediscounts. Yet, the central bank maintains that what is lacking is voluntary savings and is opposed to unnecessary raising of the levels of financing. Other pressures coming from the private sector include requests for the elimination of taxes and the granting of subsidies (Gómez, 1964). Economic stability fosters a higher volume of investment and exchange stability is an indispensable element in economic progress.

Ortiz Mena, Finance Minister, also mentions the achievements attained through stability. Saving has experienced a marked growth, for it has been duly channeled by banking institutions toward the kind of investment required by development. Institutional saving was not the only savings source to grow, for the securities market has expanded also. There was an increase in the volume of securities and with it, a slight decrease in the rate of interest. The government considers that domestic savings are the best means of accelerating development (Ortiz Mena, 1963). The objectives of monetary and financial policy, then, are to enhance the mobility of domestic resources and prevent any inflationary tendency.

VI. *Fiscal Policy*

In the post–war period, fiscal policy has been characterized by the predominance of public spending, in view of the goals of development, as well as by traditionalist features of taxation maintained until the present decade. Public spending increased because of the pressure of developmental requirements that called for sustained public investment. Adhering to its doctrine of the State as a promoter of development, the government undertook the social overhead capital projects that were characteristic of the postwar period.

Indirect taxes still play an important role in the structure of taxation, giving it a regressive bias. The tax adjustments of 1955, 1962 and 1965 led to a lesser dependence on taxes on foreign trade; indirect taxes were reduced while direct taxes were increased, and an attempt was made to foster the formation of capital. Actually the income tax reforms did not alter the general structure of incentives for the capitalization of firms. Three distinct periods, more or less characterized by tax reforms, will be examined here in order to present the fiscal policy of the postwar years. The first of these periods covers from 1945 until the reforms of 1955; the next, up to the reforms of 1962, and the last period, from the 1965 reforms up to date.

Tax Changes Up to 1955

In examining the tax reforms of 1955, and after having suggested some guidelines for future investment, Urquidi (1956) attached great importance to the facility of capitalizing reserves, which fosters investment by allowing a firm to increase its capital. Capitalizing reserves also allows banks to turn to other credit requirements, discouraging the distribution of excessive profits, and consequently, conspicuous consumption. Other aspects of the reforms of 1955 dealt with depreciation allowances, which facilitated the rapid amortization of equipment. The same author also considers that the deductions permitted for donations and cultural appropriations lend support to education and research. Nevertheless, Mrs. Navarerete (1957) expressed the belief that the formation of capital attained by the private sector could have been achieved to the same extent without a preferential system of taxation, except for tariffs. She arrived at this conclusion because of the characteristics of the tax system itself, and the inflationary process. Among other things, she mentioned that the level of taxes was very low, that there were numerous ways of evading taxes, that taxes were paid only on a part of income, and that the entrepreneur could easily pay taxes because of inflated profits (1957, p. 38). Urquidi also has doubts about the system of exemptions to new and necessary industries, for they are not significant in the determination of industrial location and they represent a loss of revenue for the public sector and benefit only a small group of firms.

Regarding the effect of fiscal policy on demand, the criticism is more severe. According to Mrs. Navarrete, the objective of a more equitable tax burden had been reached only in a very limited way: since efforts had been dedicated to the collection of funds, they had not embraced the highest goal of Mexican policy, economic development with social justice (1957, p. 38). She also came to the conclusion that the present system of taxation was regressive. In agreement with this criticism, Urquidi added that the income tax did not fall on the combined income of the individual and that it should be reformed, for it perpetuated an undesirable distribution of income and fostered the production and importing of luxury consumer goods (1956).

Both economists recommended a

broader, more progressive tax system to promote the financing of public investment. Mrs. Navarrete demanded a redistribution of the tax burden in accordance with the economic capacity of the tax payers. She also recommended that voluntary private savings be stimulated by direct financial measures and not indirectly through the absence of a personal income tax (1957, pp. 41–42). Urquidi also insisted on the need for greater tax revenues in order to meet the requirements of public investment and welfare services, such as schools, public health and so forth. According to him, the economic development of Mexico should rest, to a great extent, on an adequate and equitable taxation of income.[6] He specifically underlines the desirable characteristics of the income tax: that it is equitable, that its effects on investment and consumption should contribute to the promotion of development, and that it represent a significant and stable source of funds for the State.

The above suggests that there are no dissident opinions. Public spending, with its goal of promoting the development of the country, is viewed favorably. (Except for the extreme liberal standpoint, which does not, which does not tolerate state intervention.) Evaluations of the tax system generally agreed, that it was inadequate for the needs of the country: it was regressive, insufficient, and had operational shortcomings. Authors like Flores de la Peña and Noyola also felt that it should stimulate a greater domestic demand. These were the first manifestations of the two currents of thought that are in disagreement about the strategy of development. One maintains that action must be taken to encourage savings and the formation of capital, while the other argues that capital formation is the result of a strong domestic market. Both points of view agree on the need to reform the tax system, even if the proposed measures are different.

The Period of Tax Change of 1962–1964

Throughout 1959–1961 the Mexican economy experienced a setback, in the form of a decrease in the rate of growth of the national product due to the stagnation of private investment. This happened during the first years of attainment of price stability. The government increased its level of investment in face of the alternative of a real recession. Deficit financing, with which government spending was still operating, can be attributed to the shortcomings of the tax system, which made the problems of financing and the smallness of the securities market more acute. In order to overcome the problem, the government made revisions—not a radical change to eliminate obsolete tax laws—having as its objective to make the tax burden more equitable by gathering some sources of income into more general basic groups. It also tried to build in more incentives for private investment. Among the main tax changes were the introduction of accelerated depreciation for new investments, the widening of the tax base and an increase in the progessivity of tax rates. The economists' comments on the changes were complimentary rather than antagonistic. Although some did so with reservations, the critics justified the reform because it was a step towards changing from the former cedular system to a global system of taxation on some incomes, with cumulative rates for incomes coming from different sources, while maintaining exemptions or special treatment for certain incomes from capital. The reservations stemmed from the reforms being viewed as insufficient to meet the needs of the country.

---

[6] Elsewhere, he added that the apparatus for revenue collection was slow and lacked coordination. And since the tax burden is not equitable—for its incidence is greatest on those engaged in productive activities—he points out a contradiction in the tax system. On the one hand, private investment is encouraged, and on the other, it is discouraged through taxation.

Concerning personal income taxes on income from remuneration of personal services or returns to capital, Mrs. Navarrete felt that many shortcomings were left, despite the reforms. There were loopholes in the way interest and returns on financial securities, as well as rental income from real estate and capital gains, were taxed. There is no system of deductions which would alleviate the tax burden on lower-income groups. The progressive aspect is very limited because there is no rate for total personal income. There was still great inequality and a good deal of discrimination in the distribution of the tax burden.

A. Cervantes Delgado (1962), in particular, maintained that there were shortcomings in the structure of the income tax, such as the breakdown of income through the cedular system of sources of income, and there was still no equitability in taxation. There was no application of the basic principle that all of the income of people must be taxed regardless of its source (1962, p. 402). According to Cervantes, the trend toward the reduction of indirect taxes was more illusory than real, due to the shortcomings of the cedular system. He asserted that the system of taxation lacked flexibility; thus, it was shown as unable to stabilize economic fluctuations in the short run. Increases in the volume of revenues had been due to administrative reasons, such as a change of tariffs coinciding with the exchange effects of the devaluations, the enforcement of new tax laws or the granting of facilities to taxpayers in arrears (1962).

Among others, Cervantes, and C. Ocaña and J. Alejo (1939a&b), were of the opinion that Mexico is one of the countries where the rate of taxation is lowest. They added that because of the undesirable distribution of income there was no reason for maintaining such a small tax burden. Yet, this situation could not be remedied by improving the cedular system; there was no alternative except to institute a single tax on income. Some of Kaldor's comments on the subject (1963) gained wide acceptance. Kaldor, who was later a consultant to the fiscal authorities in the implementation of the reforms of 1965, felt that the fact that only a small part of the gross national product was captured by taxation was due, not to poverty, but to the extent of inequitability in the distribution of income (the high concentration of wealth), as well as to the regressive nature of indirect taxes. All of this, in summary, was a reflection of the inability to tax effectively the most prosperous sectors of the community. Furthermore, in practice, many direct taxes were not enforced —which can be seen by comparing the data on national income with income declared for tax purposes. There were also ample legal means for tax evasion through exemptions and omissions. A case at hand is that of the anomymous ownership of securities (more often than not these are made out to the bearer), which is an obstacle to the collection of taxes on property.

In general terms, the criticisms were aimed at the fact that the owners of capital remained in a veritably privileged tax position. Yet taxes on labor income meant that employees and workers bore the largest tax burden, besides which the collection of these taxes is the most effective because tax withholding is practiced. These ideas were held by, among others, Ocaña and Alejo who added that the basic breakdown of incomes prevented the application of higher tax rates; that even though the rates were higher now, a redistribution of income had not been possible, and that the effects on wages and salaries had hampered the development of a domestic market (1963b, p. 173). Mrs. Navarrete also argued that the tax bias in favor of the owners of capital, besides its inequitable nature, had a disadvantageous effect on the objective of development, because the owners of capital are stimulated to invest in securities with fixed returns and low taxes, instead of taking risks in new enterprises.

The Tax Reforms of 1965

The reforms of 1962 were considered insufficient. They were the object of much criticism and there were many recommendations that influenced the changes that took place in 1965.

According to Mrs. Navarrete, there are three relevant basic principles: that all incomes should be taxed, regardless of their source; that labor's ability to pay taxes is less than that of capital, so that a system of tax deductions should be worked out in order to tax in accordance with the ability to pay; and that a policy should be formulated to lend progressiveness to the tax system. Cervantes also indicated that the fiscal system should be used to supplement private saving, and that taxes should be applied with the objective of stimulating or discouraging given activities, according to their social value (1962, p. 406).

Among others, Mrs. Navarrete (1964) and Ocaña and Alejo (1963a) stressed once again the need to make fiscal policy fulfill three objectives—stimulate production activities (supply), improve the distribution of income (demand), and collect revenues without jeopardizing monetary stability or the balance of payments. Mrs. Navarrete advocated a favorable treatment for those who carry out productive activities or risk capital (1964).[7] C. Ocaña and J. Alejo (59) added that it is necessary to avoid the transference of taxes by shifting from the high-income groups to the low-income groups, since the former both demand and supply goods and services, while the latter demand consumer goods and supply personal services. Thus, Mrs. Navarrete was concerned with the promotion of output, and the other two concentrated on the distribution of income.

Among other things, Kaldor suggested that it was necessary to effect fiscal reforms for two reasons—the insufficient level of public revenues and the fact that the inequitable distribution of income could lead to political problems. He recommended that the family be considered as the basic economic unit for tax purposes. He also suggested that corporate profits be taxed at a single rate of 40 percent and, as a general criterion, that a mixed system be applied that taxed income as well as wealth, pointing out that he would consider this more just and equitable because it approached the capacity for payment of the individual. Hoyo D'Addona of the Secretariat of Finance disagreed with Kaldor on the form that direct corporate taxation should take. He felt that it would be better to maintain the principle of progressiveness. The justification for his argument was based on the fact that the structure of corporations in Mexico is such that a few people could amass large profits because of the limited distribution of shares among the public. Hoyo D'Addona added that there was no pressing need for an overall personal income tax, although he admitted the necessity of changing the cedular system of taxation (1968).

With respect to income taxation, it was said time and again that it should be applied to total income. The authors mentioned above clamored for the consolidation of the cedular system: the tax to be applied to personal income would act as a single rate, based on a concept of income that would cover all of the earnings of the individual. Kaldor, for example, called for two changes in the rates of taxation: higher levels of tax-exempt incomes and higher tax rates that would be more progressive (1964, p. 25). The same idea had been propounded by Ocaña and Alejo in 1963 (b).

---

[7] In other words, she did not believe it expedient to collect high taxes from firms, which are the organizations that carry out productive activities and investment; what should be taxed is the income of entrepreneurs.

Kaldor's preference for a tax on both income and wealth was shared by Mrs. Navarrete. A tax exclusively on income does not reflect the economic capacity of the individual, which also depends on his property assets (1964). According to Kaldor, the tax on the profits of a firm would be a means of taxing the income of its owners, and he recommended a tax exemption on undistributed profits. The essential concept of income as discussed by him includes personal income as well as corporate income. As far as personal income is concerned, after insisting that there was a bias in favor of the owners of capital, Kaldor suggested that the tax on income from wages and salaries be made annual and that generated by capital be modified in order to tax capital gains. Regarding corporate taxes, he felt that all capital gains should be taxed, that deductions should be permitted for losses (which could be reabsorbed by later profits) and for depreciation (wear and tear), that tax exemptions for "new and necessary" industries be abolished, and that, instead, direct subsidies be granted for fixed capital expenditures. Among other changes recommended by Kaldor were to abolish the distinction between non-commercial and commercial transactions; to do away with the anonymous character of the ownership of bonds, shares and obligations, and to introduce a tax on wealth, together with an overall tax on income (1964, p. 33).

Some authors regarded the policy on the administration of state property as a supplement to fiscal policy, a source of revenue and an instrument of development. Cervantes (1962), demanded that the state administer the country's resources in a more economical and uniform way, be they untransferable assets in use, or assets in the form of investment assets.

The reform of 1965 brought about an innovation that had repeatedly been requested—the transformation of the cedular system of taxation into an overall income tax. Nevertheless, opinions about fiscal policy have not changed very much. The problems are alleged to remain the same, in general terms. Those engaged in the study of these problems are concerned mainly with the formation of capital and the redistribution of income. Let us consider some comments of Hoyo D'Addona and Ortiz Mena made in 1968 and 1966, respectively.

Hoyo D'Addona, who examined the effects of fiscal policy on supply, argues that in order to promote economic development, fiscal policy should stress the use of public expenditures. Understating the time-lag between investment and its effects, especially because of the long gestation periods of the social overhead capital projects executed by public investment, he asserts that public spending is not inflationary because it is offset by the production of goods and services. Similarly, according to the same author, the concept of deficit can vary, depending on whether there is a preponderance of investment expenditures or of current spending on government consumption. He implies that the former are advantageous for economic development, although he prefers that public spending be financed by means of taxes. All of this is in agreement with a previous statement, in which he claimed that in Mexico, there is still an excess of potential demand in relation to total supply (1966). Therefore, it develops that this author emphasizes the idea that fiscal policy should seek, above all, the formation of capital in the country. He provides an argument for encouraging the reinvestment of profits by means of a tax exemption for undistributed profits. For the same reason, he advocates the simplification of the tax schedule for the small taxpayer—a tax preference in favor of the small and medium sized firm—in order to attain a harmonious and balanced development.

The Secretary of Finance, Ortiz Mena, holds a similar view: fiscal policy should reenforce saving and investment. The most suitable instruments of that policy—in terms of private investment—are the tax system and the strengthening of the securities market. Similar goals should be pursued by tariff policy. Reflecting the recommendation of Kaldor among others, Ortiz Mena maintains that the structure of the tax on the income of firms should be so designed that it fosters the financing of capital formation by such devices as exemptions for undistributed profits and accelerated depreciation. With regard to the effects on domestic demand, Ortiz Mena points out some aspects that already have been mentioned, for example, that the redistribution of income and the enlargement of the domestic market are the established objectives for fiscal policy (1966). He notes that the progressiveness of the income tax and the different treatment of income stemming from work and from capital have been the main preoccupations of the authorities. Since a greater priority is given to domestic savings, dividends and returns on fixed-income securities were excluded from the clause on personal accumulation in the Income Laws of 1965 and 1966. He stated that the developing countries are subject to structural and institutional obstacles that curtail the effectiveness of fiscal policy, which has as its goals the creation and long-term expansion of the economic and social overhead capital by means of public investment, monetary and price stability in the short run, the stimulation and appropriate channeling of private investment, the protection and strengthening of the balance of payments, and the redistribution of the purchasing power of the population (1966, p. 9).

VII. *The External Sector*

Foreign Trade

Between 1945 and 1956, agricultural output displaced mining as the sector with the greatest value of exports. Since 1957, and especially in the present decade, the rate of growth of the agricultural sector and its exports has declined considerably. At the same time the current account deficit has steadily widened during the period of price stability.

The thesis about the external disequilibrium of Mexico, or the deterioration of the terms of trade, is frequently associated with the theory expounded by ECLA on the instability of external demand. The theory covers two aspects: the long-run trend of prices and the short-run fluctuations in the volume and value of exports. In general, the ECLA thesis on the deterioration of the terms of trade was accepted by the majority of Mexican economists, who recommended different solutions.

For Torres Gaitán, as well as many others, the decreases in the rate of growth of exports in the period 1950–1960 can be explained only by the weakening of external demand, rather than by internal factors, because of such unfavorable phenomena as the decline in prices which made the reduction of export taxes and the use of compensated trade operations ineffective (1961). In order to combat this instability, there must be international economic cooperation. World market conditions are very different today from what they were when Great Britain was the creditor center of world trade. Now that this role has been assumed by the United States, it is being performed by a nation in which foreign trade represents a very small proportion of national income. Therefore, the United States has become a monopolist as a seller and a monopsonist as buyer (1954). Faced with this situation, Latin America turns out to be a marginal seller which does not affect the international position of that country. Torres Gaitán adds, however, that the trade policy of the United States makes the instability of demand more pronounced, for when the prices of raw materials decline, the United States argues that this is a market problem, but when there are price increases, the United States finds appropriate mechanisms to contain the rise of prices. Finally, international economic cooperation is in conflict with a vicious circle related to the fact that external financing is available if there is a development program. But such a program is not possible under conditions of an unstable foreign demand and the fluctuations of external prices (1954). In contrast García Reynoso (1963) maintains that there are no unsurmountable external obstacles to expansion arising from the different situations that occur in the markets for raw materials and those for manufactured goods, and that there is a tendency toward the expansion of trade in manufactures. Moreover, there are large markets that have not been exploited because of a lack of promotion and of sales technicians, especially with trading blocks such as the Latin American Free Trade Association and the European Common Market.

According to ECLA (United Nations, 1957) the solution of the problem depends on internal factors, such as the development of demand and substitution for imports. In other words, it implicitly accepts the conclusion that conditions in the international market cannot change favorably and that increasing use should be made of a protectionist policy.

Torres Gaitán was of the opinion that between 1950 and 1960 the rate of growth of imports was similar to that of national income, from which he inferred that the average propensity to import was stable. The growth of imports was parallel to that of national income, despite the increase in the price of imported goods, as well as higher import taxes and greater import restrictions. Moreover, imports expanded at a higher rate than exports to the United States. On the basis of these observations, Torres Gaitán concluded that the demand for imports is inelastic with respect to changes in external prices —an indication of the inflexibility of domestic supply—for a good part of total demand is oriented toward the foreign sector. He explains the deficit on current account in terms of the productive apparatus. According to him, there are a number or reasons for this. One of them is the higher rate of growth of income in Mexico than in the United States. Another is a constant propensity to import due to the inelasticity of domestic supply, a factor that acts together with the obstacles to exports. A third is the difference in the levels of productivity in the two countries which tends to make for a greater volume of Mexican purchases in the United States (1961).

Izquierdo provides other observations relative to the composition of imports. He regards the predominance of capital goods within imports as a vulnerable point in the economy of the country, because of the dependence of national industry on foreign sources of supply for equipment. He also points out that purchases by Mexicans in American border cities really constitute imports, even though they are counted as Mexican tourist expenditures. The large proportion of consumption goods included under this category somewhat changes the composition of imports (1964).

Some general conclusions can be deduced from the characteristics of foreign trade. Torres Gaitán, for example, points out that Mexico has had to finance its foreign trade with capital provided by the private sector and international organizations. He also infers that there is a hidden deficit in the balance of payments which is deferred by means of trade restrictions, although to date, no attempt has been made to correct it (1962). Mexico should increase its payments capacity in order to balance its financial relation at higher levels of output and trade, as required by development (1954). At this point, a change can be detected in Torres Gaitán's thought. In 1954, he asserted that there was no possibility of growth on the basis of an increase in exports greater than that of population, for the prospects of an increase in international demand were nil. Thus, he concluded that development must be based on industrialization, although, he added, it would be more practical to increase foreign exchange earnings, rather than to substitute for imports or receive foreign economic aid. A glance at his views on exports, written eight years later, reveals a different view with regard to the outlook for exports, and the same thing is true about his opinion of foreign aid. In 1954, he did not believe in it, but by 1962, he considered it decisive to finance development.

The proposed solution to the deficit on current account still falls within the realm of a protectionist policy. No economist seems to be in favor of a more liberal import policy to facilitate the exportation of manufactured goods.

Torres Gaitán describes the general objectives of Mexican trade policy. One of them has been to reach the maximum rate of development compatible with an equilibrium of the balance of payments. Another has been to increase the volume of exports with the highest degree of local manufacture. Others include limiting imports to the level of the capacity for foreign payments and improving the terms of trade through the diversification of markets and products (1962). Little has been written about the results attained by trade policy, although in general, the opinions expressed have been favorable. It has been pointed out that two tools have been used repeatedly—tariffs and import licences or quantitative controls. Of the former, it has been alleged that the average incidence of tariffs is very low—in fact it is even said to be one of the lowest in the world—although the arithmetic average used does not constitute an adequate criterion to evaluate effective protection. Another characteristic of Mexican trade policy which prevailed for several years was the decision to abstain from signing international trade agreements, such as the General Agreement on Tariffs and Trade, which might limit the country's freedom of action. However, this policy has changed to a certain extent since the Treaty of Montevideo was signed, creating the Latin America Free Trade Association.

A viewpoint held in common by such authors as Torres Gaitán (1954), Navarrete (1957), and Uriquidi (1962), is that they consider that the United States, which is a creditor in international trade, should abolish trade restrictions, while the less developed countries—and in Latin America, the countries affiliated with the Latin American Free Trade Association—should maintain their present restrictions against the industrialized countries. The justification for this position lies in the concept of trading power. In the creditor countries, foreign trade plays such a small role within the gross national product, that the elimination of trade barriers would not have a significant effect on their level of economic activity. On the other hand, the abolition of trade restrictions in the underdeveloped countries does not increase the purchasing power of the debtor countries, which actually depends on their exports. Far from producing a decline in foreign trade, the requested concessions would have the opposite effect, for the development of the underdeveloped nations represents an increased income, and consequently, a greater volume of trade. The thesis is an attempt to explain the role of foreign trade within the economy of a developing country and to introduce an appraisal of protectionism. Nevertheless, it should be noted that according to its defenders, protectionism is valid only in the trade relations between a creditor nation and its debtor, less-developed trading partners, that is, it will not help to orient the trade relations among the latter.

Tourism

There is virtually no disagreement about the beneficial role of tourism within the Mexican economy. Two important functions have been attributed to it. The first is that it has been important in financing development, for it has been a source of foreign exchange that has paid for more than one-third of all imports. Secondly, it has been an offsetting factor in the downward trend of export prices, maintaining the capacity to import (E. Perez Lopez, 1961).

Nevertheless, Torres Gaitán (1954, 1961) remarks that the rate of growth of tourist expenditures in Mexico and that of tourist expenditures abroad constitutes a potential problem. Mexican tourism abroad is increasing at a faster rate than the inflow of foreign tourists, which might accentuate the growth of the foreign debt.

Latin American Economic Integration

The topic of Latin American economic integration was of great interest toward the end of the decade of the 1950s, at which time many works were written on the subject. The motives that led to the objective of integration, besides the example set by the European Common Market, can be classified as external and internal, although the former were perhaps more important. In general terms, the predominant external factor was the lack of confidence in exports, the fluctuations in the terms of trade and the smallness of the domestic market. The most frequently repeated argument to justify and promote the integration of Latin America has been the need to continue the process of industrialization, once substitution for imports has become more difficult.

The first topic to be discussed below will be the framework of international trade within which the country's development is taking place. Secondly, the process of integration will be considered, from the time Mexico joined the Latin American Free Trade Association (1962) through the mid-1960s. Finally, opinions on the prospects of integration and relevant recommendations will be presented.

It is a generally agreed fact that the share of Latin American exports within international trade has declined. On the other hand, as a result of industrialization, capital goods represent a higher proportion of imports, which increases the rigidity of imports and the dependence of the region on other countries in terms of production goods, which, as O. Campos Salas says, are more difficult to substitute for (1959). R. Vernon had pointed out earlier that the lag in the exports of Mexico and other countries was due to the decline in the prices of raw materials (1962, p. 528).

Urquidi upholds the point of view that protectionism is asymmetrical. Thus, the protection of agricultural products of the developed countries greatly affects the underdeveloped nations, while the protection granted by the latter to domestic production of the manufactures exported by the former does not affect the developed nations (1962, p. 422). The most self-sufficient industrialized countries tend to bypass the underdeveloped countries. The underdeveloped nations, however, faced by an increasingly weak demand for their products, are unable to procure for themselves the capital goods they need. Because of this, Urquidi claims that world trade changes its structure and practices, and that the trend is toward greater protection, albeit among trading blocks. In effect, there has been a weakening of the General Agreement on Tariffs and Trade (GATT). To Urquidi, integration constitutes a means of attaining greater stability or more commercial security. On the other hand, it would also permit a better use of productive capacity, a problem that is more evident in the underdeveloped than in the developed countries, such as those of the European Common Market (1962).

According to Campos Salas (1959), almost all consumer goods are produced in Latin America, but in insufficient quantities. Hence, intraregional trade is an important tool for the creation of new market stimuli. In this way, integration would lead to an accelerated substitution for imports which would leave foreign exchange free for the acquisition of capital goods. The integration of Latin America, then is sought because of its effect in diverting trade.

Vernon (1962), who took part in a seminar on the subject, disagrees with the above proposition. He admits that world trade has changed and that the emerging nations need more stability, but the trend should be not to force import substitution but to reduce protection. He points out that GATT has made the reduction of tariffs possible, so that at present there are very few tariffs on manufactured goods, even though the opposite is true of agricultural commodities, for every country has wanted to raise the income of its rural sector. Even so, for the developing countries that have taken part in GATT, this has meant a means of obtaining new markets without reducing tariffs. But, he concludes, for Latin America and Mexico, the initial efforts at expanding exports have consisted in approaching partially protected markets, such as those of the Latin American Free Trade Association (LAFTA) (1962, pp. 528–529). LAFTA has brought about greater competition and a certain liberalization of Latin American trade among non–traditional industries. Wionczek (1966a) infers from this that there is a need to improve the present patterns of industrialization for Mexico and to think of them in terms of the entire region. While at the beginning businessmen were opposed to the liberalization of trade, today they are its staunchest supporters. Wionzcek wonders whether this is the result of the modernization of Latin America or of the need for new markets. He remarks, however, that every national authority acts unilaterally and generally outside the integration program, and that since the Treaty of Montevideo there have been no advances regarding the problems of international payments that result from trade.[8] Wionczek states that the slowness of the negotiations is not due exclusively to economic factors. The process of integration in Latin America is hampered by the disparity of the levels of development, because the industrialization programs entail an element of prestige—a motive that was overlooked in Europe. Furthermore, LAFTA does not have a mechanism for the industrialization of the less developed of the member countries. The repercussions of industrialization on the level of employment could be damaging, especially since the prevalent situation is one of underemployment; similarly, the State as such, is fairly inefficient. The barriers to the expansion of growth are greater than in western Europe. And finally, tariffs, which were a protectionist measure at first, have become a source of revenue. Economic integration is also subject to the trade and aid policies of the developed countries. LAFTA has not yet reached the point of no return; political decisions are required to make it progress, but the larger Latin American countries have shown a limited willingness (1966a). Thus, there is also a political aspect which Wionczek regards as decisive for the establishment of concrete measures and that is the willingness and determination of the governments to participate actively in the process of integration, with a view to the long run. The coordination of the domestic policies of every country, Wionczek repeats insistently, must supplement that which is asked of the developed nations with regard to their attitudes toward LAFTA (1966b).

---

[8] Wionczek is also of the opinion that the lack of progress in monetary cooperation stems from the many interpretations of the Treaty, for the monetary authorities are absorbed in the short–run problems, the plans are ambitious, there is a lack of experience, and many refuse to define their position.

Other writers, more concerned with concrete changes, consider that the only solution lies in a process of gradual transformation—a period of transition—in order to reach a definite economic integration. According to Campos Salas, the period of gradual changes should take place under a Board of Directors and an Executive Committee of the Latin American Common Market, and he recommended a flexible schedule for 10 or 15 years. Among the most important items would be annual negotiations for tariff reductions and the unification of trade policy; equal treatment of investment coming from third countries, and a program to avoid double taxation (1959). A mechanism of this sort would be propitious for the adaptation of countries with different degrees of industralization to a system that would avoid the competition of non–homogeneous regions. As a general rule, the process would stress the principle of reciprocity. Countries would enter and remain in the Latin American Common Market only if they obtained an equitable part of the benefits (Campos Salas, 1959).

Generally speaking, the essential justification of Latin American economic integration is that regional industrialization would overcome the smallness of the market. In accordance with the present situation, the prevalent idea is that integration would be a most important stimulus for industrial expansion. This current of opinion is related to concepts mentioned above concerning the stagnation of exports of primary products and the difficulties involved in expanding those of manufactured goods.[9] On the other hand, there are the shared fears about the insufficiency of the domestic market. In sum, the basis of the idea of integration lies in the enlargement of domestic markets. Campos Salas comments on the phenomenon by saying that the new markets and free competition could lead to new poles of growth and the transmission of development to those places with appropriate conditions in terms of production factors, markets and transportation. Another advantage would be that regional output would absorb many raw materials that are exported today, with a subsequent positive effect on the world prices of such commodities—making them higher and more stable—and on the economic stability of the countries of the area (Campos Salas, 1959).

On the other hand, imports from third countries would not decline because of the increase in income that would result from industrialization, and a more rational use would be made of foreign exchange. Campos Salas also foresees another important consequence of the rationalization of industrial production, and that is that plants would acquire optimum technological dimensions and economies of scale would bring a decrease in the costs of production. To this should be added a reduction in unutilized capacity, which is now sizable in certain consumer goods industries (1959). Wionczek introduces other considerations about the future performance of the process of integration when he calls for the improvement of the agreements on complementary industrialization and the execution of technical and financial assistence programs (1966b). A final point concerns the treatment of foreign investment: both feel that the establishment of basic guidelines is necessary. According to Campos Salas, this would prevent the advantages of integration from being absorbed by large international cartels. Thus, there would be a unification of policy to control dumping and subsidies to exports.

---

[9] In speaking of the outlook for Mexican foreign trade, Vernon summarizes the general consensus about exports. It is necessary to be aware that the expansion of exports of raw materials will not be fostered by favorable conditions in the near future, which means that all efforts should be oriented toward the exports of manufactured goods (1962).

With regard to tariffs, the authors are in agreement when they call for the adoption of a common tariff and identical commercial and tax regulations (Campos Salas, 1959; Wionczek, 1966b).

The unification of monetary policy is another topic that has not been neglected. Uniformity of policies with respect to price stability, exchange rates, foreign trade credits, industrial development and wages should be promoted by the governments of member countries. From this point of view, Campos Salas is in favor of aid from the developed to the less developed countries by means of commercial credits and industrial promotion (1959).

In 1963, Wionczek called for giving priority to a regional monetary mechanism, because any rudimentary clearinghouse system would provide a saving of foreign exchange. In this way, it would be possible to achieve greater ease in payments, an improvement in the utilization of foreign exchange, and the creation of an institutional framework of permanent cooperation advancing toward a system for the settlement of multilateral payments (1966b).

VIII. *Economic Development*

While economic development has been discussed extensively, there are only a few fundamental ideas on the subject, accompanied by numerous digressions which take the form of small variations on central themes. On the other hand, it is necessary to recapitulate concepts which have already been mentioned, in order to make a more or less comprehensive presentation of the topic.

It can be said, in broad terms, that there are two predominant currents of thought about the process of development. There are those who insist on considering supply as the predominant factor for the whole economic system and others who consider the distribution and consumption factors as the motivating forces. Mrs. Navarrete, while supporting the former view, also takes the latter elements into account. Thus, she states that changes in the amounts and the combination of productive factors are the prime motivating force in the process. The changes which have direct effects are the changes in technology, the capital stock and the productivity of labor (1955, p. 231). Acknowledging that each developing country cannot be considered as an isolated system, Mrs. Navarrete claims that inward growth is subject to some limitations that arise from international relations, since the accumulation of capital is only possible with imported investment goods. Furthermore, the technology—which is also imported—is not always the most appropriate. Her recommendations about economic policy echo the ideas expressed previously; in order to sustain an endogenous growth, economic policy should be oriented toward the application of technology to production, the increase of the stock of capital goods, and the increased productivity of labor (1955, p. 232).

The other school, which will only be mentioned in passing, is one already seen in the sections on domestic demand and the distribution of income. Its thesis could be summarized as that the impetus for development can arise only by means of the redistribution of income and the subsequent increase of the domestic market. The thesis is frequently laced with support for state intervention for the achievement of development. The viewpoints that are most frequently repeated are liberal, Marxist or of a third kind, which calls for a mixed economy. The latter is the predominant one among Mexican economists. In Mexico, state intervention is generally looked upon favorably. It is said that it is based on the principle of collective well-being. Emilio Mugica (1963) cites the Constitution of 1917 as the origin of the responsibility of the State to bring about structural changes, mainly through public investment. He believes that State intervention has been compatible with private enterprise and that there is no single, rigid criterion for the determination of the functions or size of public institutions. This position supports the pragmatic and flexible approach that has been characteristic of Mexican economic policy. More specifically, Mrs. Navarrete recommends the intervention of the State in key sectors which are complementary to each other and where intervention would permit a rapid increase of income, a principle which is akin to the theories of Rosenstein–Rodan and Hirschman on unbalanced development in strategic sectors. In the case of Mexico, she notes, this is true in certain export crops and mass consumption industries, such as textiles and foodstuffs. Similarly, she is of the opinion that the State must intervene as an entrepreneur–innovator and as a banker. Nevertheless, as a capital market is formed, public financing will be gradually replaced by private financing, even though there are still fields that have to depend on government financing (1955, p. 233).

Besides adopting a similar position on State intervention within the process of development, leftist writers adopt criteria of an international political nature. Carmona (1963) claims for example, that underdevelopment is not the result of a scarcity of capital, but of the political and economic subordination imposed by colonialism.

The financing of development is a subject that has concerned many authors. Among them, Sáenz is a clear exponent of the traditional current. According to him, the main thing is to decide on the amount of resources that will be earmarked for capital formation and subsequently how its cost will be distributed. The income of the country goes into consumption or investment; for this reason, consumption can be increased only to the extent that there are reductions in investment in development programs or that use is made of the monetary reserves and the holdings of foreign exchange (1958, p. 624). Moreover, in the short run, it is necessary to take into account the inelasticity in the supply of many production goods, and that wage increases without a concurrent and proportional increase in productivity are far from being a permanent solution to the social problem of the rise in prices. They would only amount to a temporary redistribution of income, in favor of some groups and to the detriment of others whose money incomes do not rise in the same proportion.[10] Economic policy as

---

[10] When there are price increases, the solution is to reduce the volume of the money stock to a level that is congruent with the amounts of goods available for consumption. If there are problems because of inflation the solution can take one of three forms: an increase in voluntary saving, forced saving because of price rises, or increases in overall taxation (Sáenz,

Sáenz points out, should seek the allocation of resources, distributing the cost equitably within the society, without affecting exchange stability. The forms that these measures can take are well known, and the most important among them occur in the tax system, price changes, stimuli for saving, credit and finance facilities for selected purposes, and changes in the creation of money by the government (1958). On the other hand, Sáenz maintains that the possibility of capital formation by means of a monetary expansion and stimulating saving is limited by the need to keep the balance of payments in equilibrium and because of the inherent contradictions between monetary expansion and the growth of voluntary saving. In fact, he holds the view that stability within development depends basically on changes in the tax system and on maintaining short–run equilibrium. He also claims that the financial and fiscal systems are inadequate for the control of price increases. He feels that there is a generally accepted assumption that the country can invest and reinvest an increasing proportion of its resources without reducing total consumption. In reality, however, price rises and devaluations have been a direct means of holding down total consumption to a level compatible with the level of actual public investment (1958, p. 630).

Navarrete (1956) considered that economic policy had been based on credit and trade policies, as well as public spending policy. But tax policy and financial policy (in the securities market) have yet to be developed. He attributes the price increases of the two decades prior to 1957 to the excess of investment in relation to saving. Excessive investment is not harmful in itself, for it is essential for developing economies. The harm usually stems from the means used to obtain it. At the time, Navarrete already felt that there had been a reduction of the excess of investment over saving in the public sector. Similarly, he favored the tax system as the source of financing for public spending.

The Process of Development in Mexico

There are divergent opinions on the form taken by the development process of Mexico. In order to organize them, a summary will be presented first of those viewpoints based on the concept of balanced growth, and in the following section an examination will be made of those that consider economic policy from the standpoint of the measures taken to implement growth.

With regard to output, the opinion held by Urquidi (1961) is of special interest, for he considers that the growth of the productive apparatus has been balanced in the past few years. The increase of output has been abetted to an important extent by technological innovation as well as social overhead capital projects executed by the public sector. Nevertheless, there have been some obstacles, such as the coefficient of capitalization, which has fluctuated between 10 and 15 percent, a rate considered by Urquidi to be insufficient for the requirements of the country. (In 1970, it was over 20 percent.) Concerning industrial growth in particular, he maintains that this phenomenon has permitted the absorption of excess rural population and an increase in the income of the urban sector. A similar opinion has been expressed by Fernandez Hurtado (1960), who also believes that the increasing output of durable goods and capital equipment is an indication of a new stage in the process of industrialization in Mexico. Urquidi still has some reservations about it. While it is true that industrialization provides a stimulus for commercial agriculture and helps to maintain the balance of payments in equilibrium through the substitution for im-

---

1958, p. 626). The second alternate would be the one to follow if the other two are not operative.

ports, the rate of industrial growth has been low. His statement is based on criteria such as external equilibrium and the ability to solve the problem of absorbing surplus labor, neither of which has been attained by development. The roots of the slowness of growth lie in the smallness of the market—a statement of dubious validity repeated continually—as well as in the lack of integration in the process of development and in the demand for imported durable consumer goods. He also considers the subsector of intermediate goods as the most vulnerable point of the industrial structure. Regarding the agricultural sector, Urquidi subscribes to structuralist concepts about the rigidity of agricultural supply, but adds that the most dynamic segments of agriculture are those which are tied to foreign markets (1961).

What solutions have been proposed? Those advanced by Urquidi are important because they encompass the opinions of other authors as well. With respect to industrialization, he recommends that decision–making be based on two criteria —external equilibrium and the prevention of unemployment. (To these, García Reynoso adds considerations of costs and location). Expansion should be vertical as well as horizontal. Industrial integration, an important objective in recent years, makes planning and its enforcement necessary in all sectors. The acceleration of the production of intermediate goods will be to the advantage of implementation of the two previous recommendations, because of their large share within exports. Moreover, the existence of a steel industry would permit an increased output of capital goods.

IX. *Attempts at Economic Planning*

In Mexico, economic planning, in a technical sense, has never gone beyond the level of well–intentioned attempts. The first attempts at planning date from 1953, when the Investment Committee, an office which was directly affiliated to the Presidency of the Republic, was created to determine the priority of different public investment projects. In 1959, the Committee became a ministerial office. After a decline in business activity in 1962, an Intersecretarial Planning Commission was created, which put into effect a Plan for Immediate Action, which defined levels of public investment and related projects. In 1965, with the evaluation of the results of this plan, another was prepared for the period 1967–1970. By 1963, a number of works had appeared which commented on the characteristics of the Planning commission, its results, its future potential, and its role in the development of the country. At the time, many opinions were expounded, which will be divided into two groups and presented below. The first group appraises the experiences of the country in terms of planning. The second examines the prospects of planning and how it could be improved.

In the first place, it is important to establish just what kind of planning has been done in Mexico. Writers on the subject agree that it has been of an indicative nature, that is, free of dictates even within the public sector. Such is the standpoint of G. Esteva (1963), and J. Zamora Batiz (1967). De la Peña, on the other hand, considers that it has been a mixture of a centralized type of planning with pluralistic characteristics, although it depends to a great extent on the ups and downs of politics. The predominant criterion has been to associate investment and development, at least, implicitly. But different approaches, such as relating productivity and development or employment and development, would also be valid (1963). This situation is not unusual, considering the prevalent current of thought among the experts on the subject, who feel that development is directly related to investment.

Some critics consider that the results

have not been altogether favorable. Using a strict definition of planning, Esteva states that it does not exist yet in Mexico. But de la Peña says only that it has not been sufficient, for long–term goals there have not been established and frequently the means are confused with the ends. According to Esteva, the well–known Plan for Immediate Action was insufficient as an economic development plan because it was formulated before a long–term plan had been completed, because it was essentially a conglomeration of individual investment projects which did not adhere to aggregate requirements, and because, despite its indicative nature, the private sector did not take part in its formulation. He believes, however, that it has been useful for future programming of public spending and in fostering an awareness of planning (1963). The two latter ideas are shared by De la Peña and Zamora Batiz. The latter mentions other results of the first steps taken in planning, such as the creation of an awareness of problems like external financing and investment in self-financing projects, which have awakened a positive interest within the private sector (1967). But these writers also point out drawbacks, some of which are linked to institutional factors. A far–reaching problem with regard to planning projects has been brought out and criticized by De la Peña (1963), who asserts that the planning project has not been realistic enough, because it does not establish structural changes in the society nor in public administration. There is ignorance with respect to the operation of the public sector. Another problem that cannot be justified, which he points out, is that the Planning Commission does not have executive functions and hence, because of a lack of authority, it becomes a mere advisory body, which detracts from the willingness to plan. The same author notes that there are errors due to false extrapolations that result from the fact that the plan does not conform to public administration conditions nor to the social structure of the country (1963).

There are some explanations of the shortcomings of or the obstacles to attempts at planning. Sharing the ideas of I. Pichardo Pagasa (1964), and J. Zamora (1967), s. De la Peña makes note of a series of factors that limit the possibilities of planning, for example, the lack of statistics, insufficient research, the scarcity of technicians, and of a lack of technical criteria for sectorial programming. He insists that frequently the means are confused with the ends, so that state intervention depends on the individuals in power. Authorities at a lower level do not take up a position that would be supplementary to the plan. And last but not least, there is no appropriate short-run planning mechanism.

The experiences under discussion have led the same authors to make recommendations for a more effective system of planning. There is agreement that the process of the formulation of the plan should be more rational and that it should not lose its indicative nature. s. De la Peña maintains that a plan should consider two distinct aspects—the "planning of planning" and the structure of the methods to be applied (1960).

X. *Economic Policy, Economists and Mexican Nationalism*

A review of the literature published on the economic policy of Mexico reveals surprising facts and many inconsistencies. (In this section, I will refer to the common denominator of published works, not to those of outstanding, well–prepared economists whose ability is unusual and who undoubtedly have been quoted in the course of this work.) Mexican economic policy has not been examined systematically. At best, it has been examined only partially, for the formulation of economic policy in Mexico is so closely tied to the goals of the Revolution and to nationalism that it is not easy to separate them. Yet, by going to the bottom of it, one can see the various interests at play, the ideological positions and the price that the community must pay for them. In this final section, general comments will be made on the analysis of economic policy reviewed in this work, some points will be touched upon that are not discussed or analyzed by Mexican economists and an attempt will be made to examine and explain the actual practice of economics in Mexico.

Three main points stand out when one reviews the literature by Mexican economists on economic policy. The first is the nearly total absence, until recently, of the analytical tools of modern mathematical economics and general economic analysis, including the field of economic policy. At times this leads to misunderstandings and confusion on controversial matters. Sometimes a concept which expresses a mathematical relationship is given a different meaning by different people due to the lack of a precise definition, or on the contrary, different terms are used for the same concept. Frequently, descriptive charts and graphs are used to illustrate hypotheses which are never proved with statistics. In fact, there is a great deal of mistrust and suspicion connected with the use of quantitative methods.

The second point that should be emphasized is the lack of interest in abstract thought. At times the object of analysis is too general and comprehensive, and that which is gained in incidental and general terms is lost in analytical shallowness of treatment of the main theme. Of course, the more variables are included in the analysis, the clearer the need for using mathematical methods to be able to handle the multiple interrelations. All this leads to a lack of precision with respect to the assumptions of the analysis, which more often than not are not made explicit. This becomes more evident when various fields are examined by an author and his assumptions inadvertently seem to change, so as to keep coherence with his stated conclusions. This obviously makes it difficult to follow the logic of the analysis and, moreover, lends itself to never-ending discussions since, as mentioned above, the merits of different works of analysis are based upon conclusions that have been derived from different assumptions, without keeping in mind the diversity of the latter.

The third point that should be brought out is that some areas have not been analyzed and yet are of substantial interest to economists in other countries. When dealt with, these points are described in a routine manner with no critical analysis. Both aspects are combined in the position of the economist in Mexico, his employment almost exclusively in the public sector, his tendency to become involved in politics, and above all his being under the influence of the nationalism which permeates Mexico. The final comments will be devoted to an explanation of the situation.

Harry G. Johnson (1967) has analyzed nationalism in recently-formed states. It is interesting to compare his observations with Mexican nationalism and with the recommendations of the economists dealt with in this paper. Johnson maintains that nationalism implies an ideological preference in economic policy for a number of goals. One of these goals is self-sufficiency; another is the entry of public property into strategic sectors of the economy, or where this would be impractical, to replace it with a broad regulation and control of private firms.

In short, nationalism contributes certain specific traits to economic policy, the most important of which is the insistence on industrialization that sometimes results in the relative neglect of agriculture and often in the deliberate exploitation of the agricultural sector in order to finance industrialization. A second characteristic that can be attributed to nationalism is the preference for economic planning. A third trait is the indiscriminate hostility towards the big international corporations. A look at Mexican nationalism is now in order.

Actually, there are two distinct phases in the economic nationalism produced by the Revolution of 1910. The first of these covers the period until 1940; the second has continued to the present day. The latter part of the first phase coincides with its most constructive period—fiscal policy was introduced as a tool to foster economic development. Defense expenditures became less important, while education acquired great importance, as did communications and irrigation projects. With the oil expropriation, foreign firms were unquestionably subjected to Mexican laws. The distribution of lands gathered force and the political power of the large landowners (*hacendados*) was abolished. However since World War II, the Government has adopted an ideology that favors economic growth, albeit in a different way. This ideology comprises all social groups and acquires a definite form—national interest, national unity, general submission to the State, xenophobia, and the casting into oblivion of the class struggle. Thus, the nationalist ideology which embraces economic development as the goal of the Revolution was established and there an attempt was made by the Party to unite all of those taking part in the economic process. For their part, the middle and upper classes ceased to be hostile to the government and began to participate in the political life of the country. At the same time, as industrialization became the main object of economic policy, gradually the government listened more and more attentively to the entrepreneurs, and public officials wielded their power by passing judgment on each individual case without applying any general pattern of criteria. Although this system undoubtedly created uncertainty and constituted a risk, the decisions were usually favorable to production, investment and profits, and it established a balance, as well as a common understanding between government officials and entrepreneurs. Thus, the dynamics of growth, imparted by the private sector, were facilitated by avoiding the stumbling blocks inherent in labor relations problems.

As a result of the industrialization policy, there was also a bias in the distribution of the national income. Profits increased by means of higher prices for manufactured goods, while the price of agricultural inputs, those supplied by the public sector and the wage levels remained stable. In effect, the policy of agricultural and industrial development led to a redistribution of income, in terms of payments to the factors of production, which favored the middle class, especially the cultured middle class and the upper class which owned the means of production, and hence represented the means for the formation of capital.

From a social point of view, the nationalism that arose from the Revolution had two extremely positive effects. The first of these was the elimination of the previous social stratification. The second was the formation of a new social and economic structure that was modern and capable of adjusting to economic development, as evidenced by greater factor mobility and by the fact that the social system can assimilate changes without destroying the mechanism of collective harmony (M. Nash, 1967). Both aspects were propitious for the continuous growth of the national product.

The nationalist policy gave employment, energy, and a purpose to a great number of dispossessed Mexicans. At the same time, it made it possible to obtain immediate returns by employing the existing factors of production more intensively, especially, because of the agrarian reform, land and labor. Similarly, demographic mobility and the distribution of land in the form of *ejidos* kept farmers from restless ferment. The increasing levels of urban employment, the paternalism of the State, and the increase in the supply of labor made the labor unions more ductile. Meanwhile, the process of economic development transferred a larger share of the increases in income toward the middle and upper classes in the form of returns to education, oligopolic profits and other payments to property.

Nationalism created a consensus of opinion in favor of the Mexican identity and simultaneously transformed the social structure and the set of values. (In the Mexican case there is also another type of national identification, namely the *indian* and *mestizo* cultural heritage.) It also played a role in economic decision-making and helped to keep the political process outside and prevent it from disturbing the economic process, with the qualification that once the ends were accepted, the means became untouchable and unquestionable. Nationalism, in the wake of patriotism and the Mexican identity, made it more and more difficult to pass judgment on political centralism, paternalism and economic policy; in effect, the public sector is immune from independent criticism. The same is true of the distortions created by economic policy. To some extent, to exalt nationalistic values is a way of preventing criticism, maintaining privileges and defending vested interests.

What has been the role of the economists in the formulation and implementation of economic policy? The formal study of economics in Mexico began in the 30's within the Law School of the National University. It was the outgrowth of the need to understand and guide the increasingly complex economic system produced by the Revolution. From the Law School it inherited the humanistic tradition which became linked to a Marxist ideology and the rejection of mathematics and quantitative methods, that it has retained until now. The other schools of economics, founded later on, have imitated the curricula of the former and are still rather small or too recent to make their presence felt yet.

The economists, like the rest of the intellectuals, are identified with the Government and are widely employed within it. At the same time, like all who take part in politics, they subscribe to all of the dicta of the Revolution. However, the very fact that they work for the Government, together with the evidence of economic development, has limited the critical approach of Mexican economists. They are frequently reduced to the justification and rationalization of economic measures that government officials and politicians have taken without consulting them. For the most part, economists, like the rest of the intellectuals, have lent their support to the economic policy that has been followed, despite the fact that it provides measures—especially with regard to imports and investment incentives—which are partial to the already considerable privileges of the industrialists. Similarly to most of the left, they share these tenets with the so-called national bourgeoisie. For example, together with the members of the National Chamber of Manufacturing Industry (*Cámara Nacional de la Industria de Transformación*), they agree with industrialization and its promotion by means of import controls and tariffs; they also agree on the position taken on limiting foreign investment. These economists also maintain some objectives in common with the Confederation of Industrial Chambers (*Confederación de Cámaras Industriales*), although not all, for they differ on foreign investment, which the Confederation considers beneficial. Such a position is logical from the standpoint of the industrialists, but not from that of the nationalistic left, which is upheld by most economists. It is well known that these economists extol a better distribution of income, which is incompatible with and contradictory to the type of inwardly–oriented industrialization, based on import substitution, import controls and oligopolistic profits that obviously distort the distribution of income. Actually, in this respect they play up to the bourgeoisie. This incongruency can be explained as a reflection of the absolute distrust of these economists for market forces and their support of economic controls. There is also a tendency for economists to support State intervention in any form, and a partiality to production. However, in the last case, the transfer of income that results from an increase in the price of manufactured goods implies a bias favorable to industrial profits.[11]

The economists have adopted these positions as part of an ideological attitude and because the leftist position is extremely attractive, in that it gives those who uphold it a sense of being endowed with a social ethic. It should be pointed out that when this position has been combined with analytical objectivity and technical competence, it has produced the best Mexican economists. But in many cases, the reason for supporting the left is the fear of the well–known epithets applied to certain attitudes. Nobody wants to be classified as a non-nationalist when that implies being a reactionary. It is much more attractive to side with the underdog and be against imperialism. It is also more convenient, in order to avoid social criticism, to adopt a nationalistic position without an objective evaluation of the means proposed for the attainment of the goals.

The economists of the extreme left, who are generally sheltered within academic work, have maintained a static position that distorts reality. It is less and less appropriate for the dynamic conditions of the country, because that outlook has not been modified in the face of the changes in the prevailing reality. A Latin American economist wrote, ". . . most of the practicing Marxists have done little more than to continue to repeat and uphold the work of the master, without trying to use his methodology to prove those areas which remained outside of his scope or to enrich the original model with the new ones introduced by historic evolution. . . . If this is a valid assertion . . . it is even more so with regard to Latin America" (Espartaco, 1964, p. 68). The Marxist influence in the formation of Mexican economists has contributed to limiting their criticism of economic policy, for the branch of economics which lends itself best for this analysis, welfare economics, applies the theory of utility—which is repudiated by most Marxists. Moreover, welfare economics has a mathematical base that is not easily understood by the ordinary, non–mathematical economist—as are most of those in Mexico—who finds it hard to grasp the collateral effects of the proposed measures. It is precisely the coherence that welfare economics gives to economic policy through the analysis of general equilibrium which permits a better understanding of the side effects, which cannot be observed with the casuistic analysis of partial equilibrium based on formal logic, practiced by these economists. Thus, it is understandable, for example, that economists support import controls and substitution for imports through domestic production in order to increase the domestic product and employment. In general, however, they do not discuss additional effects, such as the creation of oligopolistic profits—of which foreign investors come to take advantage—that distort the distribution of income in favor of the proprietor class. Moreover, they do not take into consideration the systematic increase in the price of inputs, which acts to the detriment of exports and makes it necessary to incur a foreign debt in order to maintain the rate of development. These results are very different from the nationalistic goals that inspired them.

---

[11] Undoubtedly, this was necessary in order to launch industrial development. Nevertheless, it has been shown that at present the industrial development of Mexico is not greater than the international norm, i.e., it has not advanced more than could be expected from a country of its size and output per capita. Moreover, between 1950 and 1960—when there was a decline in industrial development by international standards. In fact, comparatively speaking, the relative share of industry within the gross national product is smaller today than it was in 1950.

In this respect, some examples might illustrate the usefulness of a revaluation of the future economic policy of the country, which, needless to say, could also be useful from a political standpoint. This implies passing judgment upon some of the dearest symbols of Mexican nationalism, analyzing their effects and going beyond paying them lip service—generally done to avoid political suicide. It also means that economic information should be disclosed to allow informed criticism, and ceasing to keep much information confidential, which practice in the final analysis acts in favor of personal interests.

Once on the road toward a market economy, in accordance with the Constitution, that is viable and dynamic, it would be self-defeating to hamper it with "revolutionary" or nationalistic attitudes which hinder factor mobility, monopolize goods and impede the transmission of stimuli arising from market incentives and economic forces. In the past, a greater factor mobility and a better utilization of resources led to higher rates of development. But even today, mobility is far from adequate for the attainment of greater efficiency and, in certain cases, the economy tends to lose flexibility rather than gain it. In agriculture, for example, the collective *ejido* proved to be more productive at first than the individual *ejido* or privately-owned farms. However, the collective form of organization threatened to create a political unit that would be difficult to control, and in order to avoid this, preference was given to the individual *ejido* (S. Eckstein, 1966). The latter form of production gave rise to a market composed of small farmers, imperfectly competitive, which divested the farmer of bargaining power vis-a-vis the jobbers and other middlemen with no limitations of property and transfer—as the *ejidatario* has—and who, free of all bonds, have evolved and become strong economically, creating a bourgeoisie in medium-sized and small cities that profits at the expense of the small farmer. Although in general the collective *ejido* seems to be preferable and it would be advantageous to expand it as much as possible and enable it to absorb new technology, the scant education and civic experience of the *ejidatarios* limit its performance. By the same token, there is a scarcity of technicians who could help them achieve an adequate organization and create mechanisms for the formation of capital. Efficiency and mobility could be supplemented more effectively than at present, if, in addition, a market for the lease of lands were permitted to develop. In an autonomous way, this market would amass lands into more efficient units which would increase output—a result that is obtained directly by means of the collective *ejido*. The producers would also have greater bargaining power in dealing with the agricultural middleman. Yet the leasing of *ejido* lands is violently attacked because it would give rise to a new form of latifundia,[12] and because it is not "revolutionary," although it would seem that what really was not revolutionary was to support the individual *ejido*. The end result has been a clandestine market for the lease of *ejido* lands which does not facilitate sufficiently the creation of more productive units or the functioning of an agricultural sector oriented to the market and an adequate allocation of productive factors.

The same thing is true when import controls, in the form of protection given to domestic producers of the same goods, restrict the competitive conditions in industry and foster excessive oligopolistic profits. A gradual, greater liberalism in trade policy would force entrepreneurs to become more efficient. It would help the more competent among them, through lower input costs and a greater demand for goods, by facilitating the exportation of goods. It would accelerate technological change. It would make the processes more labor–intensive and less capital–intensive. It would lower consumer prices and allow the consumer to share the increases in productivity. It would also raise the real income of wage–earners, farmers and the middle class. However, in this case, the desire to have an industrial base as large as possible has led government officials, in some instances, to associate the people's welfare with the number of import commodities that require an import license.

In the case of natural resources, the traditional standpoint views the sale of the products of the subsoil almost in the same light as the sale of national territory. This prejudice helps to explain the decline of mineral exports in the recent period of exchange stability.

The desire to restrict commercial relations with the rest of the world to limited exports while proceeding with import substitution has led to the establishment of plants that are below optimal size, and that have operating costs that are far above the prices on the world market. Despite the large size and comprehensiveness of the public sector, economists have not written about its magnitude, operations or efficiency except to express conventional compliments. One would think that the use of natural resources and the possibility of substitution for them is conditioned by the available technology—scientific progress and its adaptation to local conditions makes for a lesser dependence on natural resources, or makes them less essential. For this reason what a national or state–owned firm does with the physical resources is not so important as what it does with human beings. In other words, the State should ensure that people are allowed to get an education, to be free and capable men, to use the available technology, to master their environment or overcome the scarcity of some resources with the abundance of others. In this respect, it is clear that the Revolution has been successful.

In order to gain a better knowledge of economic policy, it is necessary to practice a constant and sincere self-criticism. A substantial and growing number of Mexican economists are returning to the country after doing graduate work abroad. New schools of economics founded on a modern basis are now graduating competent economists. Both types are well trained in modern techniques of economic analysis and are beginning to destroy old myths with an iconoclastic attitude. Moreover in the school of economics of the National University, former students of graduate schools of foreign universities are influencing students and reforming programs, exerting pressures so that the teaching of economics will be brought up to date and, generally, questioning the old and rigid educational systems that have constrained greater improvement within the profession.

---

[12] Although it is evident that a contemporary commercial agricultural unit that is large, productive and efficient is very different from the old Porfirian latifundium, they are often taken to be synonymous. But this tends to happen when a static view is taken of economic phenomena.

References

A. Bueno Y. Urquidi, "La Iniciativa Privada y el Estado en la Promoción del Desarrollo Económico," *Actividad Económica en Latinoamerica,* June 15, 1962, 23, 5–8.

J. Campillo Sainz, "Tesis de la CONCAMIN sobre Inversiones Extranjeras," *Actividad Económica en Latinoamerica,* July, 1966, 72, 13–15.

O. Campos Salas, "Comercio Interlatinoamericano: Integración Regional," *Investigación Económica, ENE,* 1959, *XIX,* 76, 589–606.

F. Carmona, "Dependencia y Subdesarrollo Económico," *Investigación Económica,* ENE, 1963, *XXIII,* 90, 385–4230.

J. L. Ceceña, "Las Inversiones Extranjeras en México, "*Investigación Económica,* ENE April-June, 1965, *1,* 98, 271–99.

A. Cervantes Delgado, "La Política Fiscal y las Reformas Impositivas de 1962, "*El Trimestre Económico,* FCE, July-Sept., 1962, *XXIX,* 115, 391–409.

J. S. De Beers, "El Peso Mexicano 1941–1949," *Problems Agrícolas e Industriales de México,* 1953, *5,* 1, 7–134.

S. De La Peña, "Hacia la Planeación del Desarrollo en México," *Comercio Exterior,* 1960, 610–13.

———, "Sobre el Presunto Proyecto de Ley," *Comercio Exterior,* Nov. 1963, 819–820.

M. A. Duran, "Condiciones y Perspectivas de la Agriculturea Mexicana," *El Trimestre Económico,* FCE, Jan.-March, 1961, *XXVIII,* 109, 52–91.

———, "La Reforma Agraria Mexicana," *Comercio Exterior,* Jun. 1968, 493–98.

S. Eckstein, *El Ejido Colectivo en México,* FCE, México 1966.

———, *El Marco Macroeconómico del Problema Agrario Mexicano,* Centro de Investigaciones Agrarias, Comité Interamericano de Desarrollo Agrícola, Documento Preliminar, México, D. F. 1968.

Espartaco, "Crítica del Modela Político Económico de la Izquierda Oficial," *El Trimestre Económico,* FCE, Jan.-March, *XXI,* 121, 67–92.

G. Esteva, "El Mito de la Planeación Económica Mexicana," *Comercio Exterior,* Nov. 1963, 822–26.

E. Fernandez Hurtado, "La Iniciativa Privada y el Estado como Promotores del Desarrollo," in José A. Iturriaga, ed., *México, 50 Años de Revolución,* FCE, Tomo I, Mexico, 1960, pp. 595–619.

R. Fernandez y Fernandez, "Una Nueva Política Agraria," *Revista de Economía,* May 15, 1954, *XVII,* 145–48.

———, "La Reforma Agraria Mexicana: Logros y Problemas Derivados," *El Trimestre Económico,* FCE, April-June 1957, *XXIV,* 2, 143–159.

———, "Notas sobre el Problema Agrario Mexicano Actual," *El Trimestre Económico,* April-June 1960, *XXVII,* 106, 203–8.

H. Flores de la Peña, "La Mecanica de la Inflación," *Investigación Económica* ENE, 1953, *XIII,* 4, 461–481.

———, "Crecimiento Demográfico, Desarrollo Agrícola y Desarrollo Económico," *Investigación Económica,* ENE, 1954, *XIV,* 9, 519–35.

———, "Agricultura Mexicana," *Comercio Exterior,* July, 1958, 376–79.

H. Flores de la Peña and A. Ferrer, "Salarios Reales y Desarrollo Económico," *El Trimestre Económico,* FCE. Oct.-Dec. 1951, *XVIII,* 72, 617–28.

E. Flores, "La Localización de la Agricultura y los Cambios del Uso de la Tierra en México," *Boletín de Estudios Especiales,* Feb. 13, 1959, *XIII,* 151, 161–74.

———, *Tratado de Economía Agrícola,* FCE, 3rd ed., Mexico. 1964.

———, "La Teoría Económica y la Tipología de la Reforma Agraria," *Comercio Exterior,* May 1967, 379–85.

P. Garcia Reynoso, "Veinticino Años de Política Mexicana de Comercio Exterior y sus Resultados," *Comercio Exterior,* July 1962, 406–09.

———, "Evolución de la Estructura de las Exportacion, Mexicanas," *Comercio Exterior,* Aug. 1963, 569–71.

———, "La Política Mexicana de Fomento Industrial," *Comercio Exterior,* Nov. 1968, 959–64.

R. Gomez, "Estabilidad y Desarrollo-El Caso de México," *Comercio Exterior,* Nov. 1964, 778–82.

———, "Evolución del Aparto Financiero Mexicano," *El Mercado de Valores,* March, 6, 1967. 10, 199–202.

H. G. Johnson, "The Ideology of Economic Policy in the New States," in *Economic Nationalism in old and New States,* Chicago, 1967, 124–42.

M. Hinojosa Ortiz, "Reflexiones sobre una Política Agraria," *Investigación Económica,* ENE, 1961, *XXI,* 82, 211–234.

———, "Posibilidades de una Reforma Fiscal en Materia de Impuestos sobre la Renta," *Revista Fiscal y Financiera,* April, 30, 1964, *XXIV,* 202, 35–50.

R. Hoyo D'Adonna, "El Sistema Fiscal en el Financiamiento Directo del Desarrollo," *Actividad Económica en Latinoamerica,* April 1968, 93, 5–6.

R. Izquiredo, "Protectionism en México," in R. Vernon ed., *Public Policy and Private Enterprise in Mexico,* Cambridge, Mass., 1964, pp. 241–89.

N. Kaldor, "¿Aprenderán a Gravar los Países Subdesarrollados?," *Comercio Exterior,* Jan. 1963, 46–8.

———, "Las Reformas al Sistema Fiscal en México," *Comparcio Exterior,* April 1964, 265–67.

G. Lira Porragas, "Desarrollo Agrícola vs. Desarrollo Industrial," *Revista de Economía,* 1964, *XXVII,* 5, 129–40.

E. Lobato Lopez, "La Políticia Monetaria Mexicana," *Investigación Económica,* ENE, 1968, *XXIII,* 72, 557–81.

F. Lopez Rosado, "La Situación Monetaria de México," *Investigación Económica,* ENE, 1953, *XIII,* 4, 441–54.

E. Mugica, "Participación del Sector Público en el Dosarrollo Económico de México," *Actividad Económica en Latinoamerica,* Oct. 1963, 39, 49–51.

M. Nash, "Economic Nationalism in Mexico," in Harry G. Johnson, ed., *Economic Nationalism in Old and New States,* Chicago 1967, pp. 71–84.

A. Navarrete, "Las Relaciones Financieras Internacionales de México," *Investigación Económica,* ENE, 1955, *XV,* 2, 179–89.

———, "Productividad, Ocupación y Desocupación en México: 1940–1965," *El Trimestre Económico,* FCE, Oct.-Dec. 1956, *XXIII,* 4, 415–33.

———, "El Sector Público en el Desarrollo Económico," *Investigación Económica,* ENE, 1957, *XVII,* 1, 43–61.

———, "El Crecimiento Económico de México y las Inversiones Extranjeras," *El Trimestre Económico,* Oct.-Dec. 1958, *XXV,* 100, 556–69.

———, "La Inversión Extranjera Directa en México," *El Mercado de Valores,* Oct. 31, 1966, 44, 1079–1183, 1104.

I. M. de Navarrete, "El Proceso de Desarrollo Económico y la Política Fiscal," *Investigación Económica,* ENE, 1955, *XV,* 2, 229–47.

———, "Política Fiscal y Desarrollo Industrial," *Comercio Exterior,* Jan. 1956, 26–8.

———, "La Política Fiscal y la Distribución del Ingreso," *Investigación Económica,* ENE, 1957, *XVII,* 1, 83–42.

———, "Notas sobre la Distribución del Ingreso Nacional en México," *Investigación Económica,* ENE, 1959, *XIX,* 73, 31–39.

———, "La Naturaleza de la Reforma Fiscal," *Comercio Exterior,* March 1962, 138–142.

———, "The Tax Structure and the Economic Development of Mexico," *Public Finance,* 1964, *XIX,* 2, 158–99.

A. Noriega Herrera, "Las Devaluaciones Monetarias de México, 1938–1954," *Investigación Económica,* ENE, 1955, *XV,* 1, 149–77.

J. Noyola, "¿Existe una Política de Precios?" *Revista de Economía,* Aug. 15, 1948, *XI,* 8, 9–10.

———, "El Desarrollo Económico y la Inflación en México y otros Países Latinoamericanos," *Investigación Económica,* ENE, 1956, *XVI,* 4, 603–6.

——— **y D. G. Lopez Rosado**, "Los Salarios Reales en México 1939–1950," *El Trimestre Económico,* FCE, April-June 1951, *XVIII,* 70, 201–09.

C. Ocaña Y J. Alejo, "I. La Carga Fiscal en México, su Evolución y su Importancia," *Comercio Exterior,* Feb. 1963a, 80–82.

——— y ———, "II. La Carga Fiscal en México: El Impuesto sobr la Renta," *Comercio Exterior,* Feb. 1963b, 169–73.

R. A. Ollervides, "La Nacionalización y la Inversión de Captales Extranjeros en la Industria," *Comercio Exterior,* July 1966, 486–91.

A. Ortiz Mena, "La Política Financiera en los últimos cinco años," *El Mercado de Valores,* Nov. 18, 1963, 46, 597–600, 605.

———, "Contenido y Avances de la Política Fiscal," *Actividad Económica en Latinoamerica,* Oct. 1966, 75, 5–18.

E. Padilla, "La Devaluatión del Peso Mexicano: Cuatro Conferrencias," *El Trimestre Económico,* FCE, Oct. 1948, *XV,* 3, 396–412.

E. Padilla, "La Dinámica de la Economía Mexicana y el Equilibrio Monetario," *El Trimestre Económico,* July-Sept. 1958, *XXV,* 99, 349–77.

P. Padilla, "Elasticidad en las Ofertas de la Producción Agrícola Total en la Républica Mexicana," *Boletín de Estudios Especiales,* July 19, 1958, *XI,* 132, 297–309.

E. Perez Lopez, "Importancia para México del Turismo Extranjero," *Comercio Exterior,* Jan. 1961, 23–5.

———, "Analisis de la Autoridad y Estructura de la Proyectada Comisión Federal de Planeación," *Comercio Exterior,* Nov. 1963, 820–822.

I. Pichardo Pagasa, "Planeación del Desarrollo y la Reforma de la Administración Pública," *Comercio Exterior,* Aug. 1964, 533–39.

F. Rosenzweig Hernandez, "Programación y Desarrollo Agropecuario en México," *Comercio Exterior,* Oct. 1963, 726–28.

J. Saenz, "Problemas Monetarios," *Comercio Exterior,* Oct. 1957, 535–38.

———, "El Principio de Lucrecio (O Algunos Aspectos Monetarios y Fiscales del Desarrollo," *El Trimestre Económico,* FCE, Oct.-Dec. 1958, XXV, 100, 621–37.

W. Sedwitz, "La Devaluación: Su Génesis y Consecuenciase" *Revista Fiscal y Financiera,* April 1957, 11–34.

B. N. Siegel, *Inflaciõn y Desarrollo: Las Experiencias de Mexico,* México 1960.

R. Stavenhagen, "Social Aspects of Agrarian Structure in México," in *The International Scene: Current Trends in the Social Sciences,* 1966, pp. 463–77.

R. Torres Gaitan, "Aspectos de la Política de Comercio Exterior Mexicana en la Décima Conferencia Interamericana," *Revista de Economía,* Vol. XVII, Num. 6, México, June 15, 1954, *XVII,* 6, 165–78.

———, "Panorama del Comercio Exterior," *Investigación Económica,* ENE, 1961, *XXI,* 84, 763–87.

———, "La Política de Comercio Exterior," *Comercio Exterior,* April 1962, 211–17.

V. L. Urquidi, "El Impuesto sobre la Renta en el Desarrollo Económico en México," *El Trimestre Económico,* FCE, Oct.-Dec. 1956, *XXIII,* 4, 424:37.

———, "Problemas Fundamentales de la Economía Mexicana," *Cuadernos Americanos,* Jan.-Feb. 1961, *CXIV,* 1, 69:103.

———, "México Ante los Mercados Comunes," *Comercio Exterior,* July 1962, 421–22.

G. R. Velasco, "Reflecciones sobre las Inversiones Extranjeras," *Revista Bancaria,* Jan.-Dec. 1955, 16–21.

R. Vernon, "México Ante los Mercados Comunes (Otro Punto de Vista)," *Comercio Exterior,* Aug. 1962, 527–529.

M. S. Wionczek, "Las Opiniones Extranjeras sobre la Devaluación," *Revista de Economía,* ENE, May 15, 1954, *XVII,* 5, 148–52.

———, "Los Bancos Centrales y los Acuerdos Regionales de Integración en América Latina," *Comercio Exterior,* Nov. 1963, 845–848.

———, "Integración Económica Regional: Los Factores no Económicos," *Comercio Exterior,* Sept. 1966a, 688–91.

———, "Apreciaciones sobre el Desastre de

Montevideo," *Comercio Exterior,* Dec. 1966b, 916–18.

———, "Nacionalismo Mexicano e Inversión Extranjera," *Comercio Exterior,* Dec. 1967, 980–85.

———, "La Transmisión de la Tecnología a los Países en Desarrollo: Proyecto de un Estudio sobre México," *Comercio Exterior,* May 1968, 404–13.

J. Zamora Batiz, "Comentarios sobre la Planificación Mexicana, *Revista de Economía,* ENE, Oct. 1967, *XXX,* 10, 313–14.

Birf y Nacional Financiera, *El Desarrollo Económico de México y su Capacidad para Absorber Capital del Exterior,* FCE, México 1953.

Revista de Economia, Editorial, May 15 1954, *XVII,* 5, 133–34.

United Nations Economic and Social Council, *External Disequilibrium in the Economic Development of Latin America: The Case of Mexico,* Vols. I and II, New York 1957.

United Nations, Economic Commission for Latin America, *Theoretical and Practical Problems of Economic Growth,* New York 1950.

Yugoslav Economic Policy in the Post–War Period: Problems, Ideas, Institutional Developments

By Branko Horvat

TABLE OF CONTENTS

INTRODUCTION 71

I. THREE ECONOMIC REFORMS 73
  Centrally Planned Economy 73
    Institutional Development 73
    Discussion 75
  Decentralization 77
    Institutional Development 77
    Discussion 79
  Self–government Socialism 82
    Institutional Development 82
    Discussion 83

II. PLANNING 87
  Four Five–Year Plans 87
  Growth and Cycles 90
  Development Policy and Methods of Planning 92
    Development Policy and Functions of Social Plans 92
    Institutional Framework 95
    Other Issues 98

III. LABOR–MANAGED ENTERPRISE 99
  Self–Management 99
  Enterprise 103
  The Ownership Controversy 106

IV. MARKET AND PRICES 108
  Price Policy 108
    Administratively Set Prices 109
    Development of the Market 110
  Distribution Policy 113
    Wages Policy 113
    Other Issues 117
  Foreign Trade Policy 119
    Background 119
    Prologue 122
  Three Steps Towards Free Trade 124
  The What–to–do–Next Controversy 127

V. MONEY, BANKING AND PUBLIC FINANCE 130
  Banking and Monetary Policy 130
    Banking for a Centrally Planned Economy 130
    Learning by Doing 132
    Banking for a Self–Government Economy 133
    Investment Financing 138
    Anti–Inflationary or (Anti–) Anticyclical Monetary Policy 141
  Public Finance and Fiscal Policy 145
    Budget for a Centrally Planned Economy 145
    Taxation Experiments 147
    Budget for a Self–Government Economy 149
    Communal Economy 153
    Fiscal Policy 157

VI. SELF–GOVERNMENT, MARKET AND SOCIALISM 158

REFERENCES 161

Yugoslav Economic Policy in the Post–War Period: Problems, Ideas, Institutional Developments

By Branko Horvat*

Introduction

Yugoslavia has been described as one country with two alphabets, three religions, four languages, five nations and six federal states called republics. One might add that the country has a population of twenty million and that it lies in the heart of the Balkans, with all that this connotes historically. For centuries the Balkans have been a meeting place of three world cultures and three powerful religions: the Catholic West, the Greek Orthodox East and the Moslem South. In terms of contemporary economic organization we may refer to the capitalist West, the centrally planned East and the undeveloped South.

All these influences have been felt. A rather turbulent life was to be expected in a country so located and having these characteristics. The present generation of Yugoslavs has experienced all three known modern economic systems: capitalism before the war, centrally planned economy after the war and self–government socialism in more recent years. The last–mentioned system is their own innovation and so far the only one of its kind in existence. The same generation has also experienced all four modern political regimes: bourgeois democracy (in the form of a constitutional monarchy and multi–party system) before the war, fascism during the war, a one–party state immediately after the war, and self–government democracy which is now in the process of being developed. It has also lived through a partisan national liberation war and a revolution. After the war a centralized kingdom was replaced by a federal republic, and in two decades the country had three constitutions. Finally, the same generation has experienced three different economic epochs: a pre–industrial stage before the war, rapid industrialization in the two decades after the war and the recently begun stage of a modern industrial economy approaching the Western European level. Before the war, 77 percent of the population were peasants and 40 percent were illiterate. A few economic indicators will suffice to indicate the economic development that has taken place since then: (see table 1)

Illiterates still constitute close to one fifth of the adult population, but at the same time with 11 university and college students per 1000 of population the country has moved close to the very top of the world list.

Such a tremendous pace of change virtually destroyed all traditions, but it also created a new one, a tradition of no tradi-

---

* Director, Institute of Economic Sciences, Belgrade. I am grateful to Helen Kramer for linguistic assistance and helpful comments.

TABLE 1

	Before the war Yugoslavia	Western Europe[a]	1968 Yugoslavia
Production per capita:			
Electric energy, KWH	80	500–1300	1000
Crude steel, kg.	17	150–300	96
Cement, kg.	60	100–190	190
Cotton yarn, kg.	1.3	5–11	5
Energy kg.	180	2100–4300	1030
Fertilizers, kg.	3	20–65	96
Sugar, kg.	5	24–47	25
Stocks per 1000 of population:			
Radio sets	9	110–200	160
Automobiles	1	17–50	20

[a] France, Germany, Sweden, United Kingdom.
Sources: *SGS-1969. U.N. Statistical Yearbook 1956.*

tion, a tradition of change. In line with that the 1958 Program of the League of Communists—the heir of the Yugoslav Communist Party—ends with the words:

> Nothing that has been created must be so sacred for us that it cannot be surpassed and cede its place to what is still more progressive, more free, more human.

In such circumstances economic discussion displayed certain unusual features which make formal presentation somewhat difficult. Until about 1960 most of the discussion was either not put on paper, or at least not published. Further, professional articles made practically no use of references. There was a feeling of a complete break with the past, and so there was nothing to be referred to. In the same period professional literature was almost completely descriptive. That was due partly to the fact that the first university departments in economics were established only after the war. It is said that 90 percent of all scientists who have ever lived, live today. As far as Yugoslav economists are concerned, this percentage is virtually 100.

The second reason for the lack of analytical literature is to be found in the fact that there was hardly any time left for analysis. Economists were busy changing organization, institutions, and policies and keeping themselves informed about all these changes. Unless one had the inclinations of an economic historian, it did not make much sense to engage in a long-term research project. Before the book came off the press, the system had already been changed. Thus for quite some time professional economists were just describing what was happening. Description always precedes analysis.

Finally, until recently attention was mainly focussed on what Yugoslav economists call the "economic system." Economic policy in the traditional sense—the use of a set of instruments to achieve desired results in a *given* framework—hardly existed. Problems encountered were generally solved by changing the institutional framework itself. For a long time, and to a certain extent even today, economic policy consisted of an endless series of reorganizations. The search for an appropriate economic system was the main preoccupation of economic policy.

After 1960 economic organization began to assume a more permanent shape and economic discussion began to take a more familiar form. Since then use has been made of references in articles, ties with the past and with the rest of the world have been established, economic debates have become frequent and lively, professional competence has increased, and a specifically Yugoslav theory of economic policy is now beginning to emerge.

I. *Three Economic Reforms*

Centrally Planned Economy

Institutional Development: Of all European countries occupied by the fascist invaders, Yugoslavia was the only one to liberate herself by her own forces. The National-Liberation War coincided with a genuine Social Revolution. This meant two things: an unbelievably high morale, the readiness to assault heavens—as a poet said—and also a hardly imaginable degree of devastation of the country. About 1.7 million people were killed in the battles, in concentration camps, by penal expeditions and by domestic quislings. One in every nine inhabitants disappeared in this way. Almost two fifths of the manufacturing industry was destroyed or seriously damaged. About three and a half out of fifteen million people were left without shelter. The loss of national wealth amounted to 17 percent of the total war damage suffered by eighteen countries represented at the Paris Reparations Conference in 1945 (Informativni priručnik, 1948, pp. 27–29). Apart from all this, the financial system of the country was in a chaotic state; divided and occupied by various aggressive neighbors, the country was left with seven kinds of currencies (German marks, Italian, liras, Hungarian pöngos, Bulgarian levas, Albanian francs, Serbian dinars and Croatian kunas).

The first task of the new government was to repair war damages as fast as possible and to organize the economy on what were considered to be socialist principles. For this purpose all available human and material resources were centralized, and with enormous efforts and great enthusiasm by 1947 the prewar output was achieved. The program of socialist reconstruction was carried out by means of legislative and political activities.

Yugoslavia was a peasant country. Peasants participated in the National Liberation War en mass. Agrarian reform, initiated already after the First World War, had never been fully implemented because of the opposition of the ruling classes. No wonder that one of the first moves of the new state was to undertake a radical agrarian reform. The land was to be given to those who tilled it. In less than three months after the end of the war a law was passed that took away the arable land in excess of 87 acres from farmers, in excess of 12 acres from nonfarmers. Big landowners lost their land without compensation. The land that was acquired in this way was distributed among poor peasants, who received about one half of the total land, to cooperatives and state farms (Dobrinčić *et al.*, 1951, pp. 53–54).

The next crucial move, undertaken in 1946, was nationalization of private capital in industry, mining, transport, banking and wholesale trade establishments. In 1948 nationalization was extended to retail trade and catering and in 1958 to houses with more than three apartments. About one half of the Yugoslav economy, outside agriculture, had been owned by foreign capital. Of the remainder, a sizable part had been owned by the Royal government which possessed coal and iron ore mines, forests and the largest agricultural estates; enjoyed a monopoly in retail trade of tobacco, salt, matches and kerosene; and was the largest wholesale trader, transporter, importer and exporter, banker, building entrepreneur and real estate owner (Bičanić, 1962a, p. 78). Since a number of private businessmen collaborated with the fascist invader and quisling governments, their property was confiscated. Those who took part in the

Resistance—and Communist Party members did that as a matter of course—very often gave away their property without asking for compensation. And, as was already noted, many business establishments were destroyed or damaged. In such circumstances complete nationalization was politically possible, was relatively easy to carry out and did not represent an excessive financial burden.

The next move was to introduce planning by a law in June, 1946. Plans were prepared by the Federal Planning Commission, responsible directly to the Federal Government.

Everything was now ready for the new Constitution which was adopted in 1946, and in which Article 15 read: "In order to protect the essential interests of the people, increase national welfare and make proper use of all economic potentials, the state directs economic life and development through a general economic plan relying on the state and cooperative sector and exercising general control over the private sector in the economy." This paragraph may be considered as both the definition and the inauguration of a specific socio–economic system, later to be known as administrative socialism or étatism.

The year 1947 brought the First Five Year Plan which was to lay the foundation for the future industrialized and developed Yugoslavia. The Plan was extremely ambitious—national income was to be doubled as compared with the pre–war level—but in the first eighteen months it was quite successfully carried out. It appeared as though the period of violent revolutionary upheavals was over and the country settled on a well defined and predictable course of economic and social development.

However, for Yugoslavia, history had always had some surprise in store. This time the surprise was more than unexpected: it was a complete shock. In the first half of 1948 Stalin accused Yugoslav Party leaders of revisionism and anti-sovietism. Yugoslavs rejected the accusation, and soon afterwards the Cominform countries launched a full scale political and economic attack. The Yugoslav Communist Party was excommunicated from the "family of brotherly parties," various treaties were abrogated unilaterally, development loans cancelled, trade with Yugoslavia amounting to about one half of her total foreign trade reduced to virtually nothing by the middle of 1949, and a complete economic boycott established.

The first reaction on the Yugoslav side was a somewhat naive but understandable attempt to prove that Stalin and others must have been misinformed, that no one questioned orthodoxy in organizing a socialist economy, that state ownership and central planning were keystones of the system. Motivated by consideration of this sort, in January 1949 the Central Committee of the Party decided to accelerate the collectivization of agriculture. Already in an income tax law, passed in August 1948, it was stated that "the rate of taxation should be such as to foster peasants' work cooperatives by means of lower taxes." A law on cooperatives, passed in June 1949, provided a legal framework for various types of cooperatives. Individual peasants were free not to join cooperatives if they chose. But by political propaganda and various administrative and financial devices, the authorities exerted strong pressure on them to join, and they did so in great numbers.

Meanwhile the organization of the economy was modeled after the Soviet pattern. The state budget absorbed the greater part of national income. The state apparatus was running the economy directly by means of ministries and directorates. By 1950 organizational development reached the stage at which the Yugoslav economy could be considered as a model of an administratively run or centrally planned economy (Milić, 1951, pp. 126–70). This was also the climax. Already in 1950 a new development set in. The following year a complete overhauling of the economic system was in full swing. And by the end of 1951, the centrally planned economy belonged to history.

Discussion: The ideas and theories that served as guidelines in organizing the Yugoslav economy immediately after the war are to be sought in pre–war discussions among Yugoslav Marxists. They followed the well known orthodox viewpoint according to which socialism meant state ownership cum central planning. Immediately after the war there was so much to to that little time was left for leisurely reflections. Besides, everything seemed pretty clear, both theoretically and practically. One could rely on Marxist literature and on the experience of the Soviet Union, the first socialist country. What mattered most in those days was fast economic growth. And the Soviet Union showed how to achieve it.

But copying the Soviet blueprint did not produce quite the results expected. Furthermore, the savage attack of the Cominform countries forced people to reconsider their ideological positions quite thoroughly. And so preconditions were created for the emergence of a Yugoslav version of socialism.

Economic discussion before 1952 was dominated by two themes: planning for fast growth and the search for an authentic socialism. Since the former theme will be dealt with in the chapter on planning, we shall focus attention here on the latter.

The older theory maintained that in socialism there would be no market and no prices. After the Revolution, Yugoslavia was going through a period of transition between capitalism and socialism. In this period commodity relationships were still necessary because of the existence of private ownership and because labor was still heterogenous (Kidrič, 1949). Boris Kidrič—a statesman[1] who was to dominate the economic thinking of the country until his premature death in 1952—maintained that only state ownership was truly socialist (1950a, p. 8), that "the state sector was the highest form of our social ownership . . . " (1950a, p. 8). The same opinion is still held by most economists in the Soviet sphere of influence. In Yugoslavia it did not survive beyond 1950. Consistent with the above reasoning was the extolling of the significance of state planning. R. Uvalić (1948, p. 20), Kidrič (1949, p. 42), S. Kraigher (1950, p. 12) and others repeated the familiar thesis of Soviet economists about planning as a fundamental law of socialist economics. A few years later this theory was to be described as a voluntarist fallacy.

Rereading of Marx and Engles showed the possibility of great confusion in interpretation. Marx and Engels wrote seldom and very little about socialism. What they wrote amounted to two groups of statements: one dealing with the organizational form of a socialist economy, the other with the essential social characteristics of a socialist system. Marx and Engels maintained that commodity relations and the market would disappear along with private ownership; there would be comprehensive planning: production and distribution would be organized without the mediating role of money. For many decades it seemed obvious that comprehensive planning meant central planning exercised by the government, and that the absence of private ownership meant state ownership. In 1950 it was discovered that Marx had never drawn the last conclusion. In fact, and here they argued about essential characteristics of socialism, Marx and Engels denounced the state, argued that it would wither away in a classless society, talked about the self-government of producers, and asserted that "... a worker is free only when he becomes the owner of his means of production." Marx's insistance on the freedom of the individual was discovered in a statement, which was later entered into the Party program, and which reads: "The old bourgeois society with its classes and class antagonism is being replaced by an association in which development of every individual is a precondition for the free development of all" (Horvat, 1969a, pp. 105–17).

---

[1] He soon became the President of the Economic Council of the Government and the Chairman of the Federal Planning Commission. He was also a member of the top Party leadership.

Far from being truly socialist, state ownership turned out to be a remnant of capitalism, characteristic of backward countries that are building socialism, and likely to generate dangerous bureaucratic deviations (Dobrinčić *et al.*, 1951, pp. 16–18). In 1950 Kidrič wrote: "State socialism represents . . . only the first and the shortest step of Socialist Revolution . . . Persisting in state (bureaucratic) socialism . . . unavoidably leads to an increase and strengthening of privileged bureaucracy as a social parasite, to a suppression . . . of socialist democracy and to a general degeneration of the system into . . . state capitalism . . . The building of socialism categorically requires the development of socialist democracy and a bold transformation of state socialism into a free association of direct producers" (Kidrič, 1950b, pp. 5–6).

Very soon a similar position was accepted by practically all Yugoslav social scientists. M. Novak wrote that to keep state ownership would mean" . . . not the abolition of the proletariat but the transformation of all people into proletarians, not the abolition of capital but its general rule in which a specific exploitation can be and necessarily will be developed" (Novak, 1955, p. 92). Approaching the problem from a different point of view, N. Pašić came to the conclusion: "In the past state intervention in the economy was erroneously identified with socialism. If this criterion were applied to the last several decades, it would bring into socialist ranks all eminent capitalist politicians of recent times, from Baldwin and Roosevelt to Hitler and de Gaulle" (Pašić, 1957, p. 11). A. Dragičević wrote: "Nationalization of means of production and planning are preconditions of socialism, but *only preconditions* and nothing more. In order to achieve "fully developed" socialism, many more additional factors are required, in the first instance a socialist development of political relations and of economic structure of the society" (1957, p. 218). Similarly, P. Kovač and Dj. Miljević observed that "state ownership and state management by themselves lead to small or no change in the position of the producer in the production process and in his right to participate in the management of the economy. . . . In the countries in which socialist Revolution was victorious, the state, instead of becoming an organ of the working people, may and does become an organ of the state and party apparatus, which rules on behalf of the working people" (1958, p. 13). R. Milić observes that "state socialism in the USSR through bureaucratic socialism develops into state capitalism. . . ." (1951, p. 21). These statements are not quite so novel as they might sound. Already half a century ago Z. Fabri, in connection with a book by Lenin, wrote: "If the state becomes an owner, we shall have state capitalism and not socialism. . . . Under state-ownership all proletarians would become workers hired by the state instead of by private capitalists. The state would be an exploiter and that means that an entire crowd of higher and lower managers and an entire bureaucracy with all its hierarchical strata would create a new ruling and exploiting class. It looks as if something similar has already been happening in Russia . . . (Stanovčić and Stojanović, 1966, p. 164).

Lately there has been a tendency to replace "state capitalism" by an emotionally more neutral term "étatism" (Stanovčić and Stajanović, 1966, pp. 328–36; Pećujlić, 1967). The most radical in this respect is S. Stojanović, a philosopher by profession: "The term étatism denotes a system based on state ownership of means of production and state management of production and other social activities. The state apparatus represents a new ruling class. As a collective owner of means of production it employs and exploits labor. The personal share of the members of the ruling class in the distribution of the surplus value is proportional to their position in the state hierarchy . . ." (1967, p. 35).

If the state is an institution alien to socialism, who is to organize the economic process? Clearly, the only available alternative is that this task be undertaken by producers themselves. Centralization as the principle of organization is to be replaced by decentralization, centrally managed economy by a self-government economy. In the middle of 1950 a law was passed by which workers' councils were created. The draft of the law was introduced to the Federal Assembly in a speech by President Tito who said: "The slogan, the factories to the workers, the land to the peasants, is not any abstract propaganda slogan, but one which has deep meaning. It contains in itself the whole program of socialist relations in production and also in regard to social property and the rights and obligations of the workers, and therefore it can be and must be realized in practice, if we really desire to build socialism" quoted according to (Bilandžić, 1967, p. 69). By 1952 the new economic system was already in operation.

Decentralization

Institutional Development: The preparation for the New Economic System—as it was called—started with the Law on Management of Government Business Enterprises and Economic Associations by Workers' Collectives enacted in July 1950, and ended with the Constitutional Law on Principles of the Social and Political System of Yugoslavia, accepted by the Federal Assembly in 1953. The New Economic System (NES) became operational in 1952. It was transitional in character and lasted until 1960. During these eight years the country achieved the highest rate of growth in the world: per capita gross national product expanded at the rate of 8.5 percent per annum, agricultural output at the rate of 8.9 percent, industrial output at the rate of 13.4 percent (Horvat, 1963; Popov, 1968, pp. 363–64).

The law postulated that workers' collectives conduct all activities of their respective enterprises through their managing organs, Workers' Councils and Managing Boards. The Workers' Council was to be elected by all employees of an enterprise in a secret ballot. J. A. Schumpeter once remarked: "Wild socializations—a term that has acquired official standing—are attempts by workmen of each plant to supersede the management and to take matters into their own hands. These are the nightmare of every responsible socialist" (1950, p. 226). Such a nightmare was now made legal and obligatory by an act of the Belgrade National Assembly. "The principle of producers' self-management"— explains E. Kardelj, a social scientist and one of the most active political leaders—"is the starting point of every socialist politics. . . . Revolution that fails to open the door to such a development inevitably must . . . stagnate in state capitalist forms and in a bureaucratic despotism" (Kardelj *et al.*, 1956, p. 17).

In 1951 the government was busy dismantling the central planning apparatus with its ministries, directorates and administratively fixed prices. The last directorates disappeared in 1952. On December 30, 1951, a Law on Planned Management of the National Economy was passed. It replaced detailed central planning of production by planning of so-called basic proportions such as the rate of accumulation and the distribution of investment. Enterprises acquired a large degree of autonomy. In 1951 there existed numerous categories of market and planned prices. This was all replaced by a single price structure which with certain exceptions was to be regulated by the market. The rate of exchange was made more realistic by devaluing the dinar six times. And so in January, 1952, the economy was ready to embark upon a new road of decentralization.

Once it was recognized that the essential features of socialism consisted in individual freedom and the autonomy of self-governing collectives, two important consequences followed. First, the political monopoly of the state and party apparatus became incompatible with the so-conceived social system. Second, in order to be really autonomous, working collectives had to have full command over the economic factors determining their position. The former consideration led to a gradual transformation of the Communist Party from a classical political party into what I called an association of political activists (Horvat, 1969a, p. 261). The process was initiated in 1952 when the Sixth Congress of the Party changed its name to the League of Communists. The latter consideration led to a market economy with, it was intended, a minimum of government intervention.

In 1952 and 1953 several laws were passed regulating the formation, operation and termination of business enterprises. The enterprises could be set up even by a group of citizens. The director was to be appointed on a competetive basis by a joint commission of the Workers' Council and the local government. Unsuccesful enterprises could go bankrupt.

In agriculture the collectivization drive had increased the number of peasants' work cooperatives, but with its compulsory deliveries, administrative controls and the rest it depressed output.

Once the idea of an all-embracing administrative state control was abandoned, it was useless to insist on collectivization in agriculture, even more so because of the poor economic results. Ž. Vidaković gives the following explanation: ". . . the massive participation of peasants in the armed phase of the Revolution and in setting up the revolutionary political power contributed to the failure of étatist-bureaucratic socialization of agriculture, since the social-politically active peasantry did not submissively accept the administrative methods of collectivization" (1967, p. 42). In 1953 the Law on Reorganization of the Peasants' Work Cooperatives made it easy for peasants to leave cooperatives and most of them used this opportunity. Those who remained were often poor peasants

TABLE 2

	Index of output	Number of work cooperatives
1930–1939	100	—
1948	103	1217
1949	103	6238
1950	75	6913
1951	106	6804
1952	75	4225
1953	106	1165
1954	94	896
1955	116	688
1956	97	561
1957	140	507
1964	170	16

(SZS, *Jugoslavija 1945–1964*, 1965, pp. 99, 111) and that meant that the remaining cooperatives would not be viable. In order to prevent this from happening and also to curb income polarization in the villages, two months later the government carried out a new agrarian reform which reduced the land maximum to 25 acres. Since before the war nearly nine tenths of all peasant farms were smaller than 25 acres anyway, the new reform did not meet with much opposition. But the harmful effects of former policy were not wiped out. In Yugoslavia there was a long tradition of agricultural cooperatives. Forced collectivization did a great deal to discredit cooperatives. Later the general agricultural cooperatives, which were administratively established and given a monopoly in village trade, also contributed to the discouragement of a genuine cooperative movement.

After all these changes the six-year-old statist constitution became grossly inappropriate, while the time was not yet ripe for a brand new constitution. The problem was solved by a Constitutional Law, passed in 1953. Its article four States: "Social ownership of means of production, the self-government of producers in the economy and the self-government of working people in the Commune, City and District represent the basis of the social and political system of the country. . . ."

As a consequence of the self-government principle, another very important innovation found its place in the Constitutional Law. It became known as the principle of the fusion of the political and economic sovereignty of the working people. The principle was implemented by creating the Council of Producers as a new house in the Assembly. The Council was composed of representatives of collectives of business enterprises.

In the following years the government was engaged primarily in perfecting the monetary and fiscal systems. Interest rates were applied and there was some experimentation with investment auctions. Commercial banks were added to the hitherto all embracing National Bank. Reserve requirements were introduced. Local governments acquired financial autonomy.

The First Five-Year Plan (1947–1951) was extended for a year, but never really completed. The period 1952–1956 was left with only annual plans. After NES was well established, the Second Five-Year Plan covering the period 1957–1961 was launched. It was carried out in less than four years.

Discussion: While the preceding period was mostly characterized by discussion of what was *not* socialism, the theoretical approach becomes more positive now. The discussion started by an exchange of opinions on the so-called Transition Period and ended with an analysis of what was to be known as self-government or associationist socialism.

Marx wrote that the revolutionary transformation of a capitalist into a communist society could not be carried out at once. Between the two socio-economic systems there must be a short transitional period, and the state of this period would be organized as a Dictatorship of the Proletariat. Marx's analysis looked plausible and in fact proved to be a good anticipation of what happened in Yugoslavia in the first two decades after the war (Horvat, 1969a, ch. II). Around 1952, and intermittently later, the main issue of the debate was whether socialism (considered to be the first of the two stages of a communist society) is to be included in or excluded from the Transition Period (Horvat, 1951; Novak, 1952 and 1955; Perović, 1953; Sirotković, 1951; Kiorač, 1951). The debate was highly scholastic, and yet the issue was of enormous practical importance. If Dictatorship of the Proletariat is interpreted as a form of political regime, and not as the class content of the government (which is what Marx had in mind), the identification of socialism with the Transition Period will produce a command society. If the political regime is democratic, but the Transition Period extended to include socialism, the development of a classless society may be endlessly delayed. The issue was resolved in an indirect way after the essential characteristics of a self-government socialism had been elaborated.

Contrasting the Old (administrative) and New (self-government) Economic System, R. Bičanić[2] summarizes the actual developments by enumerating differences in goals, agents and means (1962a, pp. 44–47). The *goals* of the Old System were to achieve socialism by means of state power, to equalize the position of workers in relation to the state-owned means of production, and to achieve the new social order for its own sake. Individual interests of producers and consumers were subordinate to impersonal and superhuman goals of the economic system, and the state apparatus, entrusted with the achievement of this goal, was in a position to exploit the population. The New System presupposes the withering away of the state and the management of socialized property by workers, and makes the personal happiness of every individual a supreme goal.

As far as the agents are concerned, in the Old System there was centralized state management by means of a hierarchically organized state apparatus. The directives were passed down the line in an authoritarian way with little or no independence of enterprises. In the New System the state apparatus cannot interfere with the business of individual enterprises, which became autonomous. Decentralization was applied not only to economic, but also to social and political life. Authoritarianism was replaced by self-government as a basic principle of economic and social organization.

The means of the two systems are contrasted by Bičanić in the following way: state ownership vs. social ownership; central planning vs. social planning; administrative allocation of goods vs. market; administrative rules vs. financial instruments; administratively fixed wages vs. free disposition of the income of the working collectives; all-embracing state budget vs. the budget of the state administration decentralized and separated from the economic operations; consumption as a residual vs. consumption as an independent factor of development; collectivization vs. business cooperation of peasants and large agricultural estates.

In the period under consideration economists began to study intensively writings in the economics of socialism, particularly those of Western authors. This literature had hitherto been virtually unknown. I. Maksimović (1958), F. Černe (1960) and B. Horvat (1964) produced extensive critical accounts of earlier economic literature. Černe attempted to provide an acceptable definition of socialism. In his view socialism is characterized by the following three elements: (1) Equal rights of members of the community as producers. This implies social ownership. Element (1) is a precondition for (2) equal rights in terms of income distribution. This in turn implies distribution according to work. Both (1) and (2) are indispensable for the realization of (3) equal rights in political life. As citizens members of the community must enjoy political—Černe talks of socialist—democracy (1960, p. 281). It appears that socialism is essentially a philosophy of egalitarianism. Černe's definition, although never explicitly quoted—references are not popular in Yugoslavia—may be considered as commanding wide agreement among economists and other social scientists.

---

[2] Bičanić completed his study early in 1961. Essentially the same comparative analysis had already been presented by M. Popović in 1952 (1952). Evidently, the system was being developed in a consistent way.

On a less abstract level, in an important article in 1953, Uvalić described the main intentions of NES (1954). In the administrative period output was expanded regardless of cost. Now fast growth was to be maintained but cost considerations had to play an important role in the determination of the structure of output. The law of value, i.e. the market, was to take care of that. But the operation of the law of value must be restricted in two important respects: income distribution and capital formation must be controlled. Otherwise, Uvalić warned, exploitation and market anarchy will reappear. These ideas were to dominate economic policy in the next decade. But clumsy bureaucratic and often incompetent controls of income distribution and capital formation were to become more and more irksome and irritating.

The relation between market and planning has become a recurrent theme in economic discussion. Usually market and planning are visualized as two different mechanisms. In the opinion of Černe the planning mechanism is to be used for long-run and general decisions, while short-run and partial decisions may be left to the market mechanism (1960, p. 11). A similar position was taken by J. Lavrač (1958). B. Jelić explores in more detail institutional arrangements necessary to harmonize the market and planning. He argues that unbalanced growth sometimes requires interventions even outside the general framework provided by the plan (1958).

By the end of the period (1958) under consideration, NES got its first theoretical rationalization in a book by the present author (Horvat, 1964). Since the socio-economic system is conceived as an association of business, political, etc., associations, I suggested that it be called Associationist Socialism. I pointed out that the old alleged incompatibility of market and planning was nothing more than an ideological fallacy. The market is just one and at that a very efficient—device of social planning. The integration of market and planning, social ownership and business autonomy of enterprises, produces a system with interesting new practical as well as theoretical features, First reactions towards this book were negative (Dragičevic, Stampar & Horvat, 1962; 1963). Insisting on consumer sovereignty was considered to represent the (negative) influence of Western welfare economics. Insisting on rigorous technical analysis was considered devoid of social content and so anti-Marxian. Insisting on market economy was considered to reflect the influence of the Western theory of free competition. The analysis of price formation, in which interest and rent played a certain well-defined role, was said to represent a bourgeois theory.

A similar critique was voiced by some socialist economists abroad. E. Mandel maintained that "there is a definite incompatibility between socialism—or, put otherwise, a classless society and a high degree of social equality and economic efficiency —and commodity production" (1967). This is so because commodity production inevitably generates social inequality and produces waste of economic resources. The reader was not told why this should be inevitable.

In this debate B. Ward came perhaps nearest to the truth. As to the method of analysis she says: "In value theory Horvat manages to produce more or less Marxian results from more or less neoclassical assumptions" (1967, p. 519). As to the substance of the theory she concludes: "Naturally enough this regime is essentially socialist; not surprisingly, it bears a more than casual resemblance to Yugoslavia. What is surprising is that it carries a more than expected measure of plausibility . . ."

(1967, p. 509). Most of the ideas developed in this 1958 book have by now been absorbed and seem self-evident. The latest reform is based on the market mechanism and the welfare of individuals as the main guiding principles.

Self-government Socialism

Institutional Development: The last phase in Yugoslav post-war socio-economic development was prepared by a series of political, economic and constitutional reforms in the period 1958–1963. This turbulent period was inaugurated by the new *Program of the League of Communists* in 1958. Here socialism is defined as: ". . . the social system based on socialized means of production in which social production is managed by associated direct producers, in which income is distributed according to the principle to each according to his work and in which, under the rule of the working class, itself being changed as a class, all social relations are gradually liberated from class antagonisms and all elements of exploitation of man by man" (*Program SkJ* 2a, 1958, p. 133). Thus the Yugoslav variant of socialism appears to imply social ownership, self-management in the economy, the absence of non-labor income and of exploitation. The term "working class," as explained a few years later by Kardelj, was to mean "all working people who are participating in the social process of labor and in socialist economic relations" (Kardelj, 1962, p. 1531).

By 1960 the second Five-Year Plan was successfully completed. The economy was booming, self-management in enterprises was already well established and the Program paved the way to an accelerated pace of changes. The new Five-Year Plan was prepared. The Society felt ready for a new important step forward. In 1961 three radical reforms were carried out. In order to increase the efficiency of the market organization and to improve the quality of goods produced, the hitherto virtually closed economy was to be made more susceptible to the influences of the world market. To achieve that, the system of multiple exchange rates was replaced by a customs tariff, the dinar was devalued, foreign trade was liberalized to a certain extent and the country became an associated member of GATT. Since developments in the field of money and banking were lagging behind the general institutional changes, an overhaul of the entire financial organization was undertaken. And finally, it seemed inappropriate for trade unions to continue to supervise wage levels and wage differentials in self-managed enterprises. And so this control was discontinued. Since then in this field, market competition has gone further than in any other modern economy. These three reforms inaugurated in 1961 the beginning of the third distinct phase of economic development.

By that time the country was institutionally ready for the new constitution which was promulgated in 1963. Explaining the aims of the constitution. Kardelj, one of its chief architects, said that it was "not only the constitution of the state but also a specific social charter which will provide the material basis, political framework and encouragement for the faster internal development of the system of social self-government and direct democracy" (1962, p. 1533). Self-management was extended to cover not only business but also non-profit organizations. It was generalized as a principle of self-government to be applied in all spheres of economic, social and political life. In order to achieve this, the Constitution invented a new institution: the work organization (radna organizacija). Whenever people associate in order to work for a living, they create a work organization and represent a work union (radna zajednica) which enjoys basic self-government rights constitutionally guaranteed. Work organizations include enterprises and other business establishments as well as educational, cultural, medical, social insurance and other public service establishments. As a consequence the "fusion principle" of the 1953 Constitutional Law was extended to cover all work unions, and the Assembly got three houses of work unions: for the economy, for education and culture and for health and social welfare.

The three reforms of 1961 were poorly prepared, partly inconsistent and badly implemented. As one might have expected, the sensitive market economy reacted violently. Everything went wrong: in one year the rate of growth of industrial output was reduced to one half of its 1960 level, imports soared, exports stagnated, wages went far ahead of productivity. The reformers, accustomed to a tardy half-administrative economy, were taken by surprise. Planners increased targets for 1962 in order to catch up with the Five-Year Plan goals—and were, of course, deeply disappointed. The recession was deepened. It became clear that the Plan would have to be abandoned. Administrators and political bodies were deeply disturbed. Conservative politicians and economists were busy explaining the failure of the market system and demanded that central planning be reintroduced.

Heavy pumping of money into the economy helped to generate recovery in the second half of 1962. In the next year the economy was back to its normal path of fast growth. The upswing continued into 1964 ending in a boom with heavy inflation and a great balance of payments deficit. The new recession brought a new reform. Throughout 1964 assemblies were busy discussing the principles of the new reform (Savezna Skupština, 1964). In the beginning of 1965 the government administration was set to work. By May, technical preparations were completed and in July the Federal Assembly enacted the package of laws inaugurating the reform (Savezna Skupština, 1965). Significantly enough, the solution of economic troubles was sought in further decentralization, perfection of self-government autonomy, development of a more competitive market and an integration into the world economy. What followed appeared to be a second, more radical and more consistent, edition of the 1961 reform. The reform started as an economic one, but very soon produced important social and political consequences. Multicentric planning could not help but produce a pluralistic society. Reform was in its essence a new stage of the revolution; so asserted V. Bakarić, president of the Croatian League of Communists (1967, p. 231). Self-government autonomy became firmly rooted in the Socialist Establishment.

Discussion: The reform of 1961—called also NES (II)—marked the beginning of a real academic discussion of economic matters. Up to that time institutional changes had been foo fast, and economists too few, so rigorous analysis and discussion had been replaced by descriptions.

The discussion started with an exchange between Uvalić and Bičanić. Uvalić reiterated his views that income distribution and capital formation could not be left to regulation by the market. So far, distribution according to work had encountered serious difficulties. The capital market, as a device for capital formation and allocation, was unacceptable because it would lead to group ownership. Social profitability and individual profitability were two different things. The individual interest of a collective was inferior and had to defer to the social interest (Uvalić, 1962). Bičanić objected that Uvalić did not distinguish clearly between what is commonly called the economic system, and the plan. The economic system (general conditions for business conduct) used to be an instrument of the plan; now the relation had been reversed. (In fact, two years later a party congress would request explicitly that the plan become an instrument of the system instead of the system being accommodated to the plan (Šefer, 1968b, p. 29). Uvalić offered no guidance as to how to replace labor and capital markets. He really implied central planning, with operational freedom being left to planners and politicians and discipline being reserved for the rest. Bičanić feels that this is unacceptable. A modern economy is essentially polycentric and not monocentric (1963a).

In December 1962 the Association of Yugoslav Economists organized a debate in Belgrade about the draft of the new constitution (Ehonomist, 1962). A number of participants—R. Davidović, M. Macura, N. Čobeljić, K. Mihajlović—argued that the role of planning was underestimated in the draft constitution. Macura explained that this was so because economic problems were approached from the point of view of an enterprise, even an individual, instead of from the point of view of the economy as a whole (*Ekonomist*, 1962, p. 462). Čobeljić thought that planned market economy would in future be replaced by market planned economy (*Ekonomist*, 1962, p. 473). Mihajlović argued that, while consumer and intermediate goods markets worked well, investment goods and capital markets were notoriously imperfect and needed strict control (*Ekonomist*, 1962, p. 500). The debate reached its climax at another meeting a month later in Zagreb.

Futher discussion was prompted by the failure of the reform. The economy sank deeply into depression (relative to the standard Yugoslav state of affairs). From the beginning of 1961 to the middle of 1962 the annual rate of growth of industrial output dropped from 12 to 4 percent. The government was alarmed and asked a group of academic economists associated with a research institute to find out what had happened. This move set a precedent in the governmental attitude towards managing economic affairs. In a few months the group produced a report, popularly called *The Yellow Book* (Horvat, 1962a). Then the second, even more important, precedent was established: the government accepted the report.

The findings of the *Yellow Book* may be summarized as follows. Inefficient planning resulted in economic instability. The structure of supply failed to match the structure of demand, there was a downward shift in long-run export trends, there was a serious lack of skilled labor force. The inherently unstable economy was exposed to the simultaneous shocks of the three poorly prepared and badly implemented partial reforms cited above. The insistence on financial discipline created a serious shortage of money with strong deflationary effects. The abolition of income control led to wild increases of wages unrelated to productivity increases. The liberalization of foreign trade emphasized the fundamental importance of economic research as a basis for economic policy and the stability of the legal and policy framework as a precondition for efficient operations of enterprises in a market setting.

In the meantime, another research institution produced an analysis of the defects of the economic system. The report became known as *The White Book* and it criticized deficient planning, an imperfect market, arbitrariness in income distribution and inconsistencies in investment decisions (Dabčević *et al.*, 1962). Both documents were discussed in a meeting jointly organized by the Association of Economists and the Federal Planning Bureau in Zagreb in January 1963 (Savjetovanje, 1963). The former planning officials and a certain number of economists with a more centralist orientation criticized the two documents. They questioned the possibility of efficient investment and a high rate of growth in a decentralized setting. They thought that the market necessarily led to a destruction of the socialist principle of of income distribution. Some of them pointed out that the classical conflict between the essentially social character of production and atomized decision-making lay at the bottom of all economic difficulties (Savjetovanje, 1963, p. 192). However, the majority of economists agreed on the necessity of further decentralization and the perfection of self-government autonomy. Since the Zagreb debate the basic principles of the development of the economic system have never been seriously questioned among Yugoslav economists.

The well know saying about doctors—the operation was successful but the patient died—might have been applied to discussions among Yugoslav economists: the causes of economic troubles had been well explained, but the reform was dead. It soon became clear that the entire experiment had to be repeated. And so it was, in 1965. The situation was rather complicated. "The casual observer is often puzzled," commented R. Bičanić. "Only a few years ago Yugoslavia was presented as an example of a country with one of the highest growth rates in the world, now the foremost aim of economic policy is to reduce investment. For more than a decade the socialist. economy struggled against bureaucratic command; now an administrative price freeze has had to be introduced. It was the first country in the world to initiate workers management in business enterprises and to abolish the wage system; now there is discussion about whether this means too much or too little democracy. . . National problems were said to have been solved; and now the country is pregnant with increased tensions among the constituent nations, tensions newly created and socialist in origin. Efforts to find solutions to all these problems are now concentrated into two words: *The Reform.*" (Bičanić, 1966, pp. 633–34).

Bičanić and Dzeba (Dzeba and Belsać, 1965) the following aims of the reforms. The immediate purpose was to combat an increasing pace of inflation; to remove the chronic deficit in the balance of payments; to reduce all sorts of subsidies (for exports, unprofitable production, etc.) drastically in order to avoid the necessity of central administrative interventions; to correct price disparities in order to establish more efficient market relations and eliminate administrative controls. These were preconditions for some longer-term measures of structural change in the economy such as: revision of growth and investment policies; putting the productivity of the economy on an internationally competitive level; liberalization of foreign trade and elimination of the balance of payments deficit; convertibility of the currency in order to open the economy and expose it to the stimulating influences of the world market. In its broader social aspects, the reform was expected to impart a de-politicization of economic decisions; double the share of enterprises in the control of national income, reducing thereby the economic power of the state; to link the level of living to that of productivity; to increase the rationality of economic decision-making. Bičanić concludes that the fundamental aim was in fact "to build a model of a socialist system for a developed country, one which will be able to stand the competition of other developed countries without the constant tutelage of government machinery" (1966, p. 643). He and Horvat (Dobrinčić *et al.*, 1951) pointed out that this model is very different from the mixed economy of the welfare state.

The aim described was to be achieved by a process which Bičanić called the four

D's: Decentralization, De-étatization, Depolitization and Democratization.

As often happens, the ideas were good but the implementation was poor. The reform was politically much better prepared than the one in 1961, but not so economically. Economically it was based on a rather naive idea of the viability of the *laisser-faire* principle. Monetary policy appeared to be practically the only available device of economic policy. In order to stabilize prices, the government applied a credit squeeze. It worked, but it also produced deflation with unemployment and stagnation. From the beginning of 1965 to the middle of 1967 the annual rate of growth of industrial output dropped from twelve percent to minus one percent. Negative growth rates had not been known since 1952. The government thought that this was unavoidable, and that the reform "in its strategic aspects" proceeded as planned. Some economists and many businessmen were alarmed. For them, developments were catastrophic and certain to produce another failure. Soon economists were to discover the existence of business cycles. Since cycles had not been known to the government—it was held as self-evident that cycles could not exist in a socialist economy—the government proceeded to frame economic policy as if the cycles did not exist. The results of such an economic policy could not be encouraging.

The discovery of cycles proceeded in stages. The successive retardations of growth, described already in the *Yellow Book*, indicated that the Yugoslav economy might have been subject to cyclical fluctuations. The research undertaken in the Institute of Economic Studies confirmed the hypothesis. (This will be discussed more fully in Chapter 5.) In the Spring of 1967, in Ljubljana the Association of Economists held a meeting dedicated to problems of stabilization (Savjetovanje, 1966). Four papers dealt explicitly with business cycles. The research institute mentioned ventured to make a forecast of the lower turning point (1967), boom (1969) and recession (1970) of the current cycle which proved to be correct up to the time these lines were written (second half of 1969).

A couple of months after the Ljubljana meeting a public debate took place. It was focussed on the theme: "Economic Science and the Economy" (Institut, 1968a). Seven economists participated. A. Bajt raised the question of the responsibility for the reform and criticized the naive view that investment generated inflation. Z. Baletić evaluated the contention that there was a conflict between politicians and economists. Ž. Mrkušić analyzed the foreign trade equilibrium. Horvat pointed out a number of mistakes contained in currently popular economic reasoning (and, consequently, in economic policy), and in a separate article, which caused a newspaper explosion of discontent, calculated the losses due to cyclical instability. The output lost appeared to amount to about forty percent of the social product. The three remaining economists supported the official view that everything was more or less all right.

In February 1968 the Institute of Economic Studies organized an all-Yugoslav conference on the current economic situation. The study prepared for this occasion (Institute 1968b) described the cyclical mechanism operating in the Yugoslav economy and made a coherent proposal for an anti-cyclical policy. This was an important step forward. The proposal insisted on a combination of monetary and fiscal policies (the latter was virtually non-existent at that time); on a combination of price and income controls; and on the importance of the interrelations between aggregate demand and investment.

By the end of the same year another feature of the unsuccessful 1961 reform was repeated: two research institutes were officially asked to assess the implementation of the reform. There was, however, an interesting difference: this demand did not come from the government but from the Central Committee of the League of Communists. Two reports were prepared: the findings were more or less the same. I quote from the report that was published (Institut, 1969). This report found that in spite of a strong deflationary policy, prices were no more stable than they were before the reform; that the Five-Year Plan was not likely to be fulfilled; that the administrative control of prices was extended over a greater percentage of output than before the reform; that the liberalization trends in foreign trade were checked and reversed; that the balance of payments deficit was expanding; that the rate of saving was decreasing; that the losses and indebtedness of firms were increasing; that the rise in labor productivity was slightly retarded; and that unemployment was increasing beyond anything known in the country in the past two decades. Elaborating its early prognosis in more detail, the Institute predicted an acceleration of growth in the first half of 1969 (to a rate some sixty percent higher than the one forecast by the Federal Planning Bureau), an inflationary pressure in the second half and the downturn of the cycle and the beginning of a new recession by the end of 1969 or in the first half of 1970. The first two forecasts proved to be correct, the last had still the status of forecast at the time these lines were written. A few months later V. Rajković undertook to analyze the unpublished papers prepared by the administration as a basis for the reform. Rajković came to the conclusion that none of the important goals had been achieved in a satisfactory way (1969/70, p. 47).

Once again the ominous question was posed: What had happened? A careful analysis of developments seems to suggest the following answer. Economic growth and institutional changes were too rapid for the government apparatus and other organs of economic policy to be able to cope with efficiently. Almost overnight a backward Balkan country reached a European standard of economic development, and an adminstrative economy was transformed into a market economy. At the same time responsible authorities often lacked the necessary understanding of how a modern market economy operated. If to all that we add the pioneering in the system of self-government—nonexistent anywhere else in the world—it becomes clear that the complexities of the socio-economic environment have increased enormously and that it will take some time before the organizational framework is adapted, the necessary knowledge is accumulated and the new social system begins to operate smoothly (Institut, 1968a, 1969; Horvat, 1968a).

II. *Planning*

Four Five-Year Plans

The rationale for central planning was explained in Chapter 1. By 1947 the machinery for central planning was completed. Hierarchically organized planning commissions on various levels—federal, state, district and city—were entrusted with comprehensive planning in their respective territories. The operational planning and implementation was carried out by ministries and then down the line by general and chief directorates, and planning sections in the enterprises. Annual plans were broken down into quarterly, monthly and ten-day plans. In 1949 about 13,000 groups of commodities were planned (Čalic, 1948, p. 15). In the same year the state budget comprised two thirds of the national income (Kidrič, 1960, p. 453). Every enterprise had to send to the superior authorities 600–800 different reports per year. The annual economic plan weighed some 3,300 pounds (Bičanić, 1957, p. 65). Supplies and customers were assigned to every enterprise in advance. Since these administrative allocations were not quite perfect, the enterprises were asked "to find their ways." The planning authorities would provide them more money then they wanted and would ask them to spend it. As prices were fixed, spending money meant finding raw materials and investment goods necessary for the fulfilment of the plan. In a market economy one endeavors to save money, in the centrally planned economy one is at great pains to spend it: in the former selling is the most difficult task, in the latter buying is the greatest worry of businessmen.

The economy was run as one single mammoth enterprise. That required establishing a system of continuous control of operations of all enterprises. In 1948 Kidrič voiced complaints against those who considered that there was no need for daily reporting and who were satisfied with ten–day reporting (Kidrič, 1960, p. 468). A number of years later J. Stanovnik, now Secretary of the U. N. Economic Commission for Europe, at a lecture delivered to Swedish economists in Stockholm, was asked what sort of devices were used to implement plans in Yugoslavia. He answered: "Telephones!."

The first Five–Year Plan covered the period from 1947 through 1951. It proclaimed four main goals:

(1) to overcome economic and technological backwardness;
(2) to strengthen the economic and military power of the country;
(3) to strengthen and develop the socialist sector of the economy;
(4) to increase the general welfare of the population.

Consumption was taken care of, but it was last in the order of priorities. The goals enumerated were to be achieved by an explosive increase of output; compared with the pre–war level, national income was to increase 1.9 times, agricultural output 1.5 times, industrial output 4.9 times. However, due to poor statistics, the pre–war level must have been greatly underestimated and the three targets were achieved only by 1954, 1959 and 1961 respectively.

At first the implementation of the plan proceeded in a satisfactory way, though not as well as was generally believed.[3] In 1949 the economic blockade of the Cominform countries forced Yugoslavia to search for trading outlets for about one half of her exports and to secure the same proportion of imports from other sources. Although substantial foreign aid was secured two years later, this sudden reorientation of foreign trade had stifling effects on growth. The next blow came from nature; in 1960 a severe drought reduced agricultural output by one third. Collectivization also helped to aggravate agricultural problems. The radical economic reorganization in 1951 could only complicate matters. Industrial output fell by four percent in 1951, and by one more percent in 1952. The plan was extended for a year, but that was already pointless, and the report on

---

[3] Thus V. Begović reports about the overfulfillment of the first half of the Five–Year Plan (1949). But later statistical estimates showed that the data produced by the Federal Planning Commission (Informativni Pritučnik pp. 251, 484) were inflated. Thus for the output of manufacturing, mining and power plants the differences are as follows:

	Indices		
	1948/1946	1949/1948	1950/1949
Federal Planning Commission	267	116.6	106.3
Federal Statistical Office (later estimates)	190	111	103

the fulfilment of the First Five-Year Plan was never published.

And yet, if not a full success, the Plan was far from being a failure. It generated output substantially above the pre-war level, it raised the share of gross investment in fixed assets to 33 percent of gross national product (material product definition: close to 30 percent on the SNA definition) and created entire new industries.

In 1952 rigid central planning was replaced by "planning by global proportions." These proportions were: minimum use of output capacity and the corresponding wage fund, profits as a percentage of the wage bill (a device for wage planning), basic capital formation, taxes and allocation of budgetary resources (Vučković, 1952, p. 31). In this way, the central plan was expected to regulate general economic activity without administrative orders, by influencing the rate of growth and the proportion between investment and consumption, and by effecting structural changes in the economy (Jelić, 1961). The old Planning Commission—which acted as a superministry controlling the activities of all economic ministries and was in charge of the overall implementation of the plan (Djordjević, 1965)—was replaced by the Federal Planning Bureau, an expert institution with no administrative powers. Republics, districts (later communes) and enterprises would produce their plans independently. State planning became social planning which meant wide consultations among all interested parties, inclusion of non-profit institutions and independence of enterprise plans.

The next three years were used to complete the key investment projects of the Five-Year Plan in annual installments. In the discussion about the 1955 Plan the new mood was already apparent; agriculture looked neglected, investment too large and onesided (Popović, 1964, pp. 147, 150). By the end of that year M. Popović could say in the Federal Assembly that one period of economic development was completed (1964, p. 160). The year 1956 was used to prepare the Second Five-Year Plan for the period 1957-1961. In this plan increase of consumption already ranks third among the five main goals (Lovrenović, 1963, p. 220). Growth of investment was somewhat retarded and its structure radically changed. The share of industrial investment was substantially reduced in order to double the share of agriculture and increase the shares of transport and trade (Popović, 1964, p. 211). Within manufacturing, consumer goods industries were to expand faster. So-called non-productive investment in social overhead capital was also accelerated. All these changes proved beneficial and the plan was carried out in four years. The planning system seemed to be well adapted to the needs of the economy and worked satisfactorily. This system was described by J. Sirotković (1961), S. Dabčević (1963), and Jelić (1962).

The first plan distorted the structure of the economy by emphasizing capital formation in heavy industries. The second one undertook to make corrections but went to the other extreme by overexpanding consumer goods industries. Thus, the third plan was left with the task of redressing the balance again by accelerating investment in power generation, metallurgy and intermediate goods industries. These fluctuations in investment induced Čobeljić and R. Stojanović to invent a theory of investment cycles inherent in a socialist economy with an uneven pace of technological progress (1966). Z. Baletić, Bajt (1969) and others criticized this theory as unacceptable since mistakes in planning are attributable to ignorance and not necessarily to socialism, and that technological progress is rather innocent in this respect.

The Third Five-Year Plan for the period

1961 through 1965 endeavored to accelerate the growth of output even further. Personal consumption ranked second among the goals (Lovrenović, 1963, p. 221). The Plan was hardly launched when the country found itself in the middle of a recession, the reasons for which were explained in the previous chapter. The Plan was doomed to fail. In order to avoid unpleasant discussions, the Federal Assembly decided to replace it by a seven-year plan covering the period 1964–1970. For that purpose the Assembly passed a Resolution in which the basic political and economic goals of the new plan were defined as follows (*Yugoslav Survey*, 1964):

(1) steady rise of the level of living, in the first place of personal consumption, and higher share of personal incomes in national income;
(2) catching up with international standards of productive efficiency and labor productivity;
(3) expansion of external trade through more intensive inclusion of Yugoslavia in the international division of labor;
(4) accelerated development of underdeveloped areas;
(5) further development of socialist soviety by strengthening the role of direct producers and working organizations in the management of productive forces.

A comparison of these goals with those of the First Five-Year Plan shows very clearly the distance that separates social planning from state planning. The welfare of individuals is moved from the end to the beginning of the priority list.[4] Behind this change one finds the philosophy which holds that economic welfare is both the purpose and the most powerful incentive for production. An autarchical orientation is replaced by openness towards the world market and international influences. The measure of the perfection of a socialist economy is no more to be found in increasing the share of the state in the national capital but in the development of self-government. Yet, the First Plan and the Resolution had one thing in common: neither of them was implemented.

The Resolution in fact foreshadowed the Reform of 1965. The changes in economic institutions were so radical that it became necessary to prepare a new Five-Year Plan for the period 1966–1970. The Plan incorporated the goals of the Resolution. It envisaged a somewhat lower rate of growth of GNP (7.5–8.5 percent per year), a relatively modest expansion of manufacturing (9–10 percent), but a high rate of productivity increase (6–7 percent a year). Current analysis of the Federal Planning Bureau indicates that these targets are not likely to be achieved (Medenica, 1968).

Growth and Cycles

In order to be able to evaluate successes and failures in planning—and in economic policy in general—one has to have a look at some data. The following table summarizes the developments in terms of rates of growth of the most important statistical aggregates.

In the central planning period collectivization caused stagnation in agriculture and the economic boycott of the Cominform countries caused stagnation in exports. As a result total output grew slowly. In the second period the unfettered economy was in full swing in all spheres with an acceleration of growth in the second half of the period. Foreign trade expanded faster than output and exports faster than imports. In the third period agricultural out-

---

[4] Personal consumption was reduced at a rate of 2 percent annually in the period 1948–1952; it began to expand at 4.6 percent per annum in 1953–1956; it expanded at approximately the same rate as national income, at 9.3 percent, in 1957–1963; and its rate of growth surpassed that of national income afterwards (Šefer, 1965, pp. 207–209).

TABLE 3.—GROWTH OF THE YUGOSLAV ECONOMY 1946–1968 (RATES OF GROWTH, PERCENT PER ANNUM)

	Central Planning 1946–1952	Decentralization 1952–1960	Self-government 1960–1968
Gross National Product	2.3[a]	9.8	6.8[c]
Industrial output	12.9	13.4	7.9
Agricultural output	−3.1[a]	8.9	2.1
Export of Commodities	−3.1[b]	11.7	7.0
Import of Commodities	3.6[b]	9.7	7.0
Employment[d]	8.3[a]	6.9	2.4

[a] 1947–1952
[b] 1948–1952
[c] 1960–1967
[d] Persons employed outside private agriculture.
Sources: *Statistical Yearbooks.*

put caught up with domestic demand, while the European export markets became increasingly difficult to penetrate. The slowly expanding market for agricultural products reduced the rate of growth of agricultural output. Increased economic instability depressed the average rate of growth of manufacturing. As a result, the overall pace of expansion was reduced. In all these developments institutional factors, described before, played an important role. If one wants to judge the performance of the economy on the basis of a somewhat longer period, the period 1952–1968 appears to be the appropriate one. In these sixteen years, total output expanded three and one half times, manufacturing five times, agriculture two and a half times, foreign trade in commodities four times and employment outside private agriculture three times.

Since Yugoslavia has been so far the only country that has lived through three different economic systems—capitalist, étatist and self-government—in a relatively short period of time, it may be possible to evaluate the comparative efficiency of the three systems. Something of the kind was attempted by T. Marschak. He reduced the dimensions of the problem by studying the comparative efficiency of the centralized and decentralized frameworks. Marschak's results were not conclusive. He felt that the lessons which the current designer of new economic systems could draw from the Yugoslav experience were "staggeringly obscure" (1968, p. 586). Later research was undertaken in the Institute of Economic Studies (IES) (Horvat, 1969b). Efficiency was measured in terms of the rate of growth of output attributable to technical progress, defined as the residual after the contributions of labor and capital have been accounted for. The results are summarized in Table 4.

The periodization in the table is not ideal and is determined by the availability of data. Yet the results of the analysis are extremely suggestive. In the foregoing section it was stated that the investment program of the First Five-Year Plan was completed by 1955. Statistical testing in the IES study showed that the Yugoslav economy operated on the basis of two completely different production functions, one applying to the period 1947–1955 and the other afterwards. The former had a negative residual, the latter a positive and a very large one. The table seems to suggest that central planning expanded output and employment fast, and capital formation even faster, as compared with the private capitalist pre-war economy. But it also reduced overall efficiency. Self-government

TABLE 4.—THE USE OF LABOR AND CAPITAL AND TECHNICAL PROGRESS IN YUGOSLAVIA

		GNP	Rates of growth per annum in %		Rates of growth of GNP due to increased efficiency
			Employment	Fixed Assets	
Capitalism:	1911–1932	3.28	1.87	3.52	0.71
	1932–1940	4.67	0.72	2.59	3.16
Etatism:	1940–1954	5.91	4.76	9.99	−1.04
Self-government:	1956–1967	10.31	4.44	7.84	4.44

Note: The war years 1914–1918 and 1941–1945 are excluded. The data refer to manufacturing, mining, power generation, construction and crafts.

accelerated the growth of output and technical progress beyond anything known before while preserving fast employment expansion.

As might have been noticed already, the growth was fast but not at all smooth. At first the possibility of regular cyclical development was rejected by some economists. Yet in another IES study business cycles with periods of about four years were established (Horvat, 1970). These cycles, that manifest themselves as fluctuations in the rates of growth (see Figure 1), have interesting features not found elsewhere. Thus inventories are accumulated in the downswings and decumulated in the upswings; the accelerator is not operative; prices tend to vary inversely with the cycle etc. The upper turning points seem to be generated by divergent changes in import and export elasticities that end in an explosion of the balance of payments deficit. The lower turning points are somewhat more difficult to explain. Bajt believes that consumer demand is to a certain extent autonomous and helps to generate an acceleration of output growth (1969b).

If the beginnings of the cycles are measured from inflection points in the downswings of the rates of growth (these points correspond to peaks of deviations from an exponential trend of absolute magnitude), they appear to coincide with major economic reforms. Thus, the five cycles that have occurred so far describe in an interesting manner the history of post-war economic policy (Roman numerals indicate quarters):

1. Cycle: New Economic System (1), III/1949–III/1955.
2. Cycle: The transition to the Second Five-Year Plan, III/1955–II/1958.
3. Cycle: New System of Income Distribution, II/1958–IV/1960.
4. Cycle: New Economic System (2), IV/1960–L/1965.
5. Economic Reform, I/1965–?

Cyclical institutional development seems also to be a novel feature of business cycles.

Development Policy and Methods of Planning

Development Policy and Functions of Social Plans: The philosophy of development, generally accepted by Yugoslav economists and the Government until about 1956, is well described by Čobeljić, then the deputy director of the Federal Planning Bureau (1959a). Čobeljić maintains that rapid industrialization is the chief method of generating development. Industrialization creates additional urban employment, which alleviates latent un-

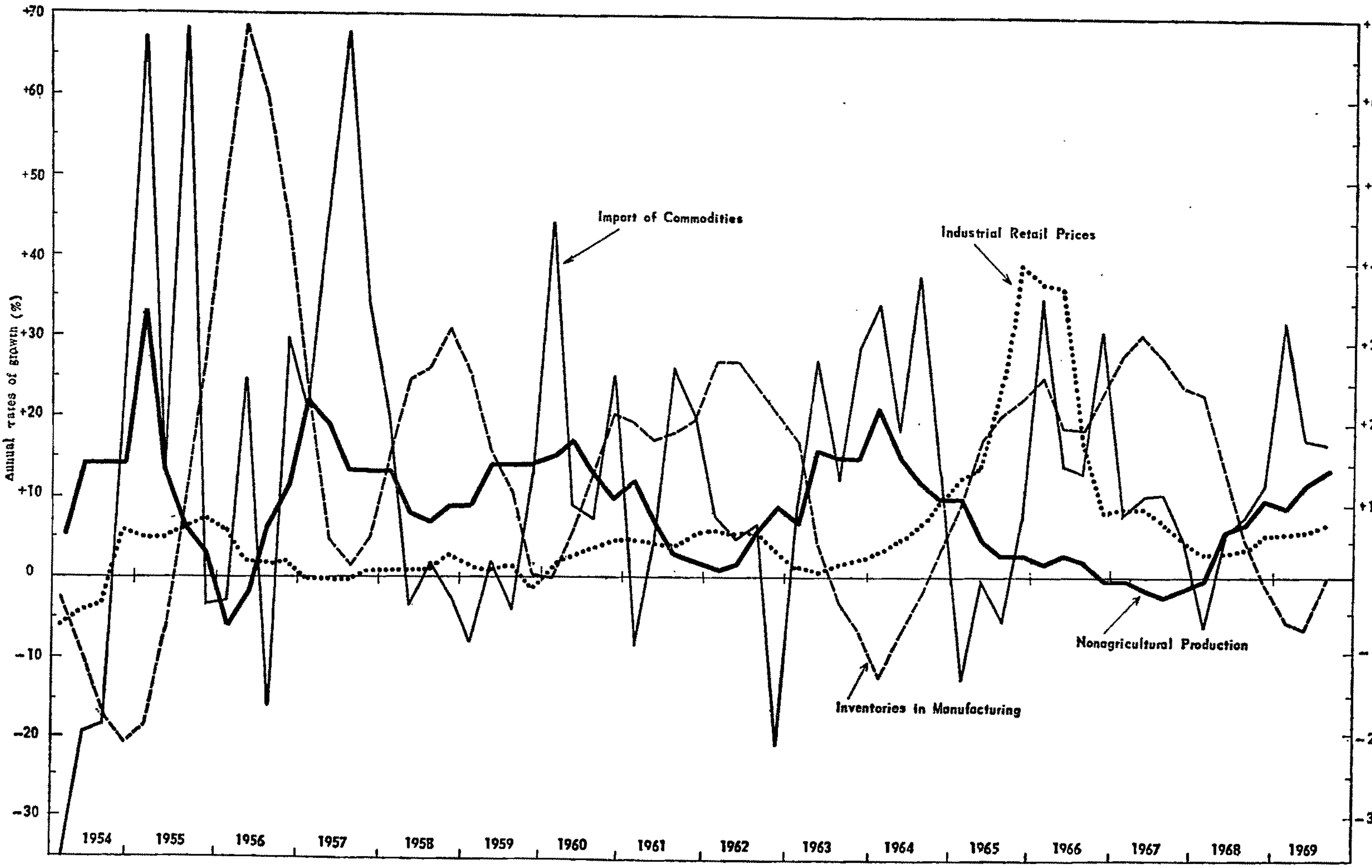


FIGURE 1. BUSINESS CYCLES IN YUGOSLAVIA

employment in agriculture. The growth of the urban labor force generates additional demand for agricultural products and so stimulates the development of agriculture. Physical control of foreign trade, in order to prevent the import of non–essential goods and to secure imports of capital goods, and the more rapid growth of consumer goods industries (so–called Department II) helps to accelerate industrial growth which in turn generates development impulses throughout the economy. Imports are paid for by exports of raw materials and agricultural products. The necessary saving is secured by a proper price policy. Prices in predominantly private agriculture are kept low and in consumer goods industries are inflated by means of high turnover taxes.

The policy described was not only advocated but was also consistently implemented. In the period 1950–1956 investment in industry (manufacturing, mining and power generation) absorbed 51 percent of all investment. The share of industry in national income rose from 21 percent in 1939 to 40 percent in 1956. Four fifths of industrial investment were channeled into heavy industry and power generation. The share of saving in national income increased four times as compared with the pre–war level (Čobeljić, 1959a, pp. 178, 366).

The planning system in 1947–1952 was consistent with such a development policy. The main characteristics of this system, as described by Jelić, another deputy director of the FPB, were as follows: a/ strict centralization of decisions about priorities, timing and structural changes; b/ physical allocation of resources as a basic method of planning; c/ financial elements play a secondary role and serve to achieve balances in value terms; d/ targets represent directives; e/ production is planned by commodities and capital formation is planned in terms of individual investment projects; f/ prices are administratively fixed; g/ the elements of which a plan is composed are also instruments of its implementation (Jelić, 1962, pp. 102–105).

After the engine of growth had been set into motion in the way described and the economy organized along socialist lines, there was a possibility of and a need for a different approach. Čobeljić now expected a more balanced growth. Jelić referred to Rostow's take–off theory and to Bićanić's threshold of growth theory (Bićanić, 1962b) and insisted on decentralized initiative as a further vehicle of growth. Self–government implied that the function of planning be separated from the function of operational management. Jelić pointed out that social plans should determine at least three global proportions—the basic division of national income, the structure of investment and the relations with foreign countries—if they were to be efficient devices for implementation of social preferences (Jelić, 1962, p. 144).

The same three global proportions were accepted as basic by D. Bjelogrlić, director of the Planning Bureau of Serbia. He added, however, a fourth one: the relative growth of the less developed states and regions (1965, p. 118). Bjelogrlić presented his paper to a conference on social planning held in Belgrade in 1965, where Čobeljić and K. Mihajlović spoke in favor of introducing more directives into planning, while M. Samardžija and M. Korač maintained that even the planning of the share of accumulation and the structure of investment meant a violation of self–government. This discussion, which covered a wide spectrum of opinions from semi–central planning to an almost complete *laisser faire* approach, has been characteristic of the Yugoslav economic profession since the enactment of the new Constitution in 1963. The trend has been towards the *laisser faire* extreme. In 1960 the Federal government controlled 48 percent of business investment directly through its General Investment Fund, and in addition to that 14 percent indirectly through tied loans (Jelić, 1962, p. 158). In 1969, the Party Congress recommended that so-called state capital be eliminated and in the future the federal government is not supposed to retain any direct control over investment resources. A satisfactory solution has not as yet been found, and work on the new Law on Social Planning, which began in 1963, has not yet been completed (Savezna Skupština, 1966a).

The advocates of the new approach to planning, Sirotković (1966), the former director of the Planning Bureau of Croatia, and R. Štajner (Savezna, 1966a), the present director general of the Federal Planning Bureau, M. Mesarić (1967) and others argue that the professional function of planning should be supplemented by an emphasized social function, that annual plans should be abandoned and replaced by parliamentary resolutions (which has been practiced since 1966), and that medium-term plans should be continually revised and extended every two to three years. Bićanić describes the desirable system of planning as polycentric planning. This presumes the existence of one planning mechanism consisting of many autonomous plans interlinked in a specific, competitive way (Bićanić, 1963b, 1967). These ideas have been more or less accepted, but in parliamentary debates criticisms have been voiced that it was not at all clear how the plans were to be implemented (Savezna Skupština, 1966a, p. 91). In practice the implementation of plans has left much to be desired and the law on Social Planning is still to be produced.

The functions of social planning in the present Yugoslav setting have been described by the IES (Jugoslavenski Institut, 1968, p. 20), and similarly by Mesarič (1969), as follows: (1) A plan is, first of all, a forecasting device. (2) As such it provides economic subjects with necessary information for their autonomous decision-making. This, together with institutionalized consultations, makes the plan an instrument of coordination of economic decisions. (3) After relevant social preferences have been determined by an essentially political process, the application of modern tools of economic policy makes the plan an instrument for programming economic development. (4) Once the social Plan has been adopted by the Parliament, it becomes a directive for the Government. Point (4) is the only administrative or compulsory aspect of social planning.

Institutional Framework: The precondition for efficient social planning is an adequate analysis of the functioning of an institutional framework. A general idea of how the system works or is supposed to work may be obtained from a description by Horvat (1969c).

The Yugoslav economic system consists of autonomous, self-governing, work organizations[5] and individual producers in market and non-market sectors and of government machinery. The task of the latter is to use *non-administrative* means in coordinating the activities of market and non-market agents and to organize public administration in certain fields of common interest (judiciary, defense, foreign affairs, etc.).

The functioning of this economic system is based on the assumptions that the self-governing collectives are materially interested in maximizing their incomes and that the Government and Parliament are able to create an economic environment in which autonomous decision-makers behave in accordance with general social interests. Both assumptions seem to have been proved correct by the modern theory of economic policy and by experience in well organized market economies. Between the "Center" (Parliament) and the "Periphery" (Work Organizations) four types of gravitational forces are active in keeping the system in equilibrium and the economic agents on the predictable trajectories of social interest.

---

[5] "Working Organization" is a constitutional term meant to underline a fundamental equality in rights and status of every group of citizens organized with an intention to earn a living regardless of the activity they perform. An enterprise, a theatre and a government office—all of them are work organizations.

These forces are information—consultation ties, market ties, economic policy ties (instruments of economic policy and legislation) and administrative ties. The last mentioned are exceptional as far as economic agents are concerned and apply to various organs of the Center such as ministries, the National Bank, certain bureaus, and the like.

I should add that there is also a fifth type of ties—political ties—which closes the whole structure connecting the work organizations with the Parliament and with flows of commands (arrows) oriented from the Periphery towards the Center. In order to keep this section short, I shall not analyze these ties (this is why they are omitted from Figure 2). It is, however, important to realize that the Parliament is organized in a rather unorthodox fashion. Apart from the traditional Political House, whose members are more or less professional politicians, elected by all citizens, there are three additional houses, dealing with three different social–economic groups of problems (economic, health and welfare, education and culture). The members of these three "Houses of Work Unions" are not professionals; they keep their usual jobs and are elected by the "producers" in these three specific fields.

Let us now have a look at the market half of our economic cosmos. The activities of enterprises and individual producers are coordinated by the market in the first place. The market is, however, a very rough and unreliable mechanism requiring constant adjustments.

These adjustments are achieved through general regulative measures and the instruments of economic policy of the Government. The financial flows, intended to achieve a desirable allocation of resources, are regulated by the National Bank within the framework of the Social Plan. There are two additional types of specific financial interventions: in the field of foreign trade (credits and exchange risks insurance) and in investment (insuring proper structure and regional allocation of capital formation). These three purposes are served by three federal funds: for export credits, for underdeveloped regions and for investment.

Market equilibrium is being worked upon by three institutions. Two of them—the Directorates for food and for industrial products reserves—intervene whenever supply and demand do not match. The former Directorate also administers agricultural support prices. The third institution, the Price Control Bureau, is now a somewhat alien element in the system. I expect that in the near future this governmental bureau will evolve into a Price and Wage Arbitrator, an institution in which all relevant interests would be represented and all decisions made jointly. At the moment more than 40 percent of industrial prices are controlled.

Statistical and Planning Bureaus have only informative–consultative functions in this system.

A rather peculiar arrangement of the Yugoslav system is to be found in what I call a *quasi–market*. The activities of schools, hospitals, museums, and other non–market work organizations cannot be coordinated by the market directly as is done in the case of enterprises. In a socialist society sick persons should be healed, talented youths educated, regardless of whether and how much they can afford to

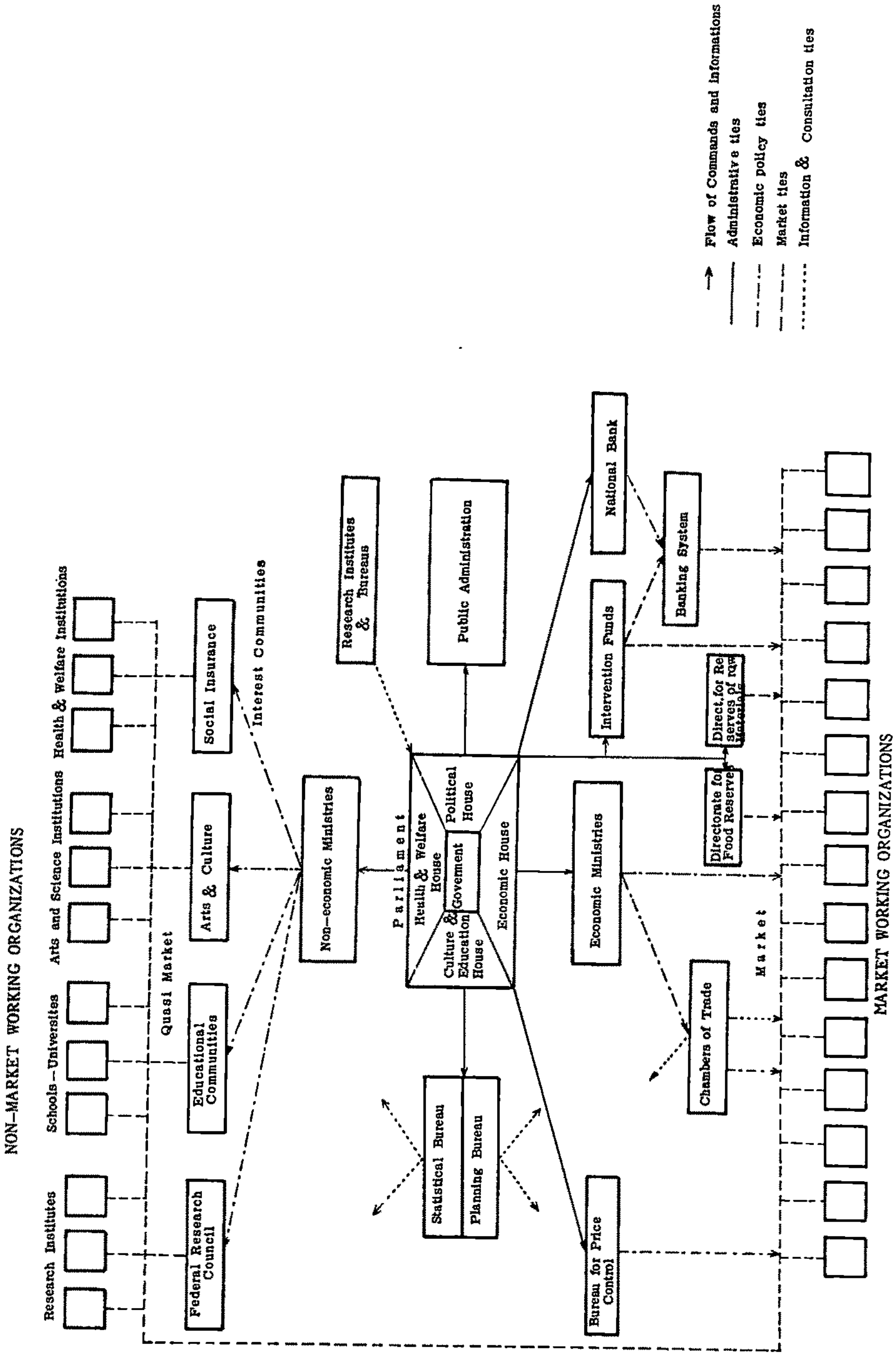


FIGURE 2. INSTITUTIONAL MODEL OF THE YUGOSLAV ECONOMIC SYSTEM

pay. On the other hand, the traditional budgetary financing of non-market activities has led to bureaucratic practices incompatible with a self-government system. The solution of this dilemma was sought in an interpolation of a special self-government mechanism between the government and the non-market working organizations. This mechanism is called interest unions. The unions obtain their financial resources on the basis of parliamentary decisions and then buy the services of non-market producers on behalf of the society. The non-market producers compete for available resources by offering their services on differential conditions. In this way, there emerges a special type of market-a quasi-market—which makes it possible for the relations between the non-market sector and the society to be economically conditioned, for the collectives in the non-market sector to preserve their self-government autonomy, and at the same time for relations within the fields of education, culture and social welfare to be based on the principle of "distribution according to needs," which is one of the preconditions of a socialist society. It is clear that the enterprises can also intervene in the non-market sector—either by buying services directly or by creating special foundations—and that is why in Figure 2 market and non-market sectors are also directly connected by market ties.

Apart from economic relations between federal bodies and economic agents there are relations between federal and state and local authorities, between the latter two and the work organizations and among all of them. I must, however, refrain from describing all these relations, although they are extremely important for the functioning of the system as a whole.

Other Issues: One of the recurring issues of the planning controversies is the problem of optimum investment and saving. Impressed by the unpleasant contraction of personal consumption during the Five-Year Plan, Čobeljić maintained that a certain minimum rate of growth of consumption represents the upper limit for the share of accumulation and for the growth of output (1959a, p. 188). Similarly, Stojanović argued that reduction of consumption growth below a certain limit reduces the growth of labor productivity and that this functional relation determines the optimum rate of investment (1960). Bajt also agreed that the optimum rate of investment is determined by the rate of consumption that maximizes the productivity of labor (1958), but this is not necessarily the socially desirable rate of investment. The purpose of production is to increase economic welfare, and a maximization of welfare through time can be ascertained only by a discounting procedure (Bajt, 1963). Horvat argued that pure time discounting is inconsistent because an individual will regret his present impatience at some later date, and utility discounting is impractical since it cannot be ascertained empirically. The other often suggested theoretical solution—social determination of the terminal stock of capital—is irrelevant, since no sensible planner ever insisted on carrying out a long-term plan. One constructs, say, twenty-year plans in order to take into account all relevant consequences of the decisions that are taken *now;* with every new element of information in time the plan is revised and the planning horizon pushed forward. The alternative approach suggested can be described as follows. Since every economy has a definite and very strictly limited capacity to absorb investment (in Yugoslavia the limit is around 35 percent of GNP, SNA definition), maximum growth is achieved when the marginal efficiency of investment is reduced to zero. If a lag of several months for achieving a certain level of consumption is an acceptable price for maximizing consumption within one's lifetime, then the maximum rate of productive investment is identical to the optimum rate of saving. Thus maximization of the rate of growth appears to be a proper target for socialist planning. The trouble with the First Five-Year plan was not a low level of consumption but an inappropriately high level of investment. Pushed into the region of negative marginal efficiency, investment depressed output. A large part of such a stagnating output (up to 20 percent of national income) was used for defense. On both counts potential consumption was seriously reduced (Horvat, 1958, 1965).

Sophistication in economic analysis and planning methods has advanced considerably since the telephone age described by Stanovnik. Yet, both are still far from being impressive. Interindustry analysis has been adapted for planning purposes (Sekulić, 1968; Horvat, 1969d). Interindustry analysis was used in calculating the new exchange rate and the price levels in the last reform. Simple econometric models are now regularly used in the early stages of the preparation of a plan (Nikolić, 1964; Horvat, 1968b). An integrated system of social accounts, specially adapted for planning needs, has been produced recently (Horvat, 1968b; 1969e). For the rest, planners rely on abundant statistics, old-fashioned balancing and hunches. A satisfactory methodology of planning is yet to be written.

III. *Labor-Managed Enterprise*

Self-Management

Self-management is undoubtedly the most characteristic of Yugoslav institutions. Further developed into social self-government, it is the pivotal institution of the Yugoslav socioeconomic system. Moreover, Yugoslav social scientists are quite unanimous in believing that without self-government socialism is impossible (Fiamengo, 1965). Thus the fate of socialism depends on the feasibility and efficiency of self-government. In this section we will be concerned only with self-government as applied to business firms, which is usually denoted as self-management.

Self-management is not a Yugoslav invention. The development of this institution can be followed from the beginning of the last century (Horvat, 1969a, ch. 5). Every social revolution from the Paris Commune onwards attempted to implement the idea of self-management. In the very beginning of the revolution in Yugoslavia, in 1941, workers were assuming control over factories in various places (Tanić, 1963, p. 30). With the establishment of central planning, the idea of self-management suffered a setback. However, already in 1949 it was revived; by the end of that year workers' councils were created as advisory bodies in 215 major enterprises and in June 1950 the law passed that inaugurated the era of self-management.

For more than a decade the basic organizational principles of self-management remained unchanged. All workers and employees of a firm constitute the work collective (radni kolektiv). The collective elects a workers' council (radnički savet) by secret ballot. The council has 15 to 120 members elected originally for one year and recently for a two-year period. The council is a policy making body and meets at intervals of one to two months. The council elects a managing board (upravni odbor) as its executive organ; the board has 3 to 11 members, three quarters of whom must be production workers. The director is the chief executive and is an ex officio member of the managing board.

As soon as it was established, self-management met with criticism and skepticism. Both came mostly from abroad. It was said that self-management would erode discipline and that workers would distribute all profits in wages, thus reducing the growth potential of the economy. In 1955 Ward suggested that workers had no real choice in the election of the council and that actions reportedly taken by the councils might represent rubber stamping (Ward, 1957; Horvat and Rašković, 1959). In evaluation of these criticisms one may point out that, regarding labor discipline, an International Labor Organization mission found in 1960 that ". . . while the self-government machinery for labor relations has curtailed the former powers of the supervisory staffs, it would not appear to have impaired their authority. . . . It has undoubtedly strengthened the position of the collective vis-à-vis the management, but it does not appear to have undermined labor discipline" (International, 1962, p. 203). As to the growth potentials, the rate of accumulation remained high with a chronic tendency towards overinvestment and with a high rate of growth. Elections are supervised by courts, and all candidates approved by the majority of the workers are included in the voting list. The safeguards against the creation of a managerial class are the workers' majority in the managing board and the provision that members of self-managing bodies may be elected only twice in succession.

The real difficulties were encountered elsewhere. The original organizational scheme proved to be too rigid, and had to be revised extensively in all its three components. It soon became evident that the director's position was not quite compatible with the new arrangement, and directors came to be "one of the most attacked and criticized professions in the country" (Novak, 1967, p. 137). In the étatist period the director was a civil servant and a government official within the enterprise. He was in charge of all affairs in the enterprise and responsible exclusively to the superior government agency. In the self-management system the director became an executive officer of the self-management bodies, while at the same time continuing to represent the so-called public interest in the enterprise. This hybrid position has been a constant source of conflicts. At first the director was appointed by government bodies. In 1952 the power of appointment of directors was vested in the commune. In 1953 public competition for the director's office was introduced and in the selection committee the representatives of the commune retained a two-thirds majority. In 1958 workers' councils achieved parity with communal authorities on the joint committees authorized to appoint and dismiss directors of the enterprises. The present state of affairs is that the director is appointed by the workers' council from candidates approved by the selection committee on the basis of public competition. He is subject to re-election every four years, but may also be dismissed by the workers' council. Since the appointment of the director does not depend exclusively on the will of the collective—as is the case with all other executives—he has been considered a representative of "alien" interests in the firm. There have been constant attempts to reduce his power, which have made his position ambivalent and reduced his operational efficiency. On the other hand, as G. Leman remarks, the director is expected to play the triple role of a local politician, a manager and an executive (1969, p. 28). In the context of what has just been said, the managing board was supposed to exercise control over the work of the director and the administration. Involved in problems of technical management and composed of nonprofessionals, the managing board often proved to be either a nuisance or ineffective. For professional management the director had to rely on the college of executive heads (kolegij), which was his advisory body and subordinated to him. Thus two fundamentally different organizational setups were mechanically fused into one system. The director's office provided a link between them, i.e. between the self-management organs and the traditional administrative hierarchy.

Finally, in any somewhat larger firm one single workers' council was not sufficient if there was to be real self-management. In 1956 workers' councils on the plant and lower levels were created apart from the central workers' council. Even this was not sufficient, because hierarchical relations between workers' councils at various levels were not compatible with the spirit of self-management. "The self-management relation in its pure form is polyarchic and not democratic"—explains D. Gorupić—"the democratic relationship represents a domination of the majority over the minority. . . . The polyarchic character of the self-management relationship is revealed in equal rights of members of a certain community" (Gorupić, 1969, p. 16).

In 1959 an interesting new development began with the creation of so called economic units (ekonomske jedinice). The enterprises were subdivided into smaller units with a score or several scores of workers. Since a year earlier the enterprises had become more or less autonomous in the internal division of income, it was thought that a strong incentive could be built into the system if economic units recorded their costs, took care of the quality of output, use and maintenance of machinery, and themselves distributed their incomes according to certain efficiency criteria. In an interesting study Lêman, a German student of Yugoslav self-management, argues that economic units resulted from endeavoring to eliminate dividing lines between three fields of activities: policy making, managing and executive work (1967, pp. 38–39). Soon, economic units began to practice collective decision making on all sorts of matters. It became advisable to enlarge economic units so as to comprise individual stages of the technological process or separate services. Economic units were transformed into work units (radne jedinice). The hierarchical self-management relations within the enterprise called for a revision. Important self-management rights (distribution of income, employment and dismissals, assignment to jobs) were transferred to work units. Direct decision making at meetings of all members of the work unit became the fundamental form of management. In this way the work unit provided a link between the primary group and social organization. It was both a well defined techno-economic unit, meeting the requirements of efficient formal coordination, and the basic cell of workers' self-government (Županov, 1962).

Work units, several workers' councils and managing boards, many commissions and committees—all this made the formal organization of a labor-managed enterprise rather complicated and inefficient. In order to make such a formal system work, it had to be simplified in practice and this was done in various informal ways. That in turn meant further limitations on competent professional management and a further reduction of efficiency. Workers' management is passing through an efficiency crisis caused by the need for a radical transformation of inherited organizational structures. After all, workers' management meant a fundamentally new principle in running enterprises and it would have been surprising if that did not require painful adaptations and deep changes in social relations. I must add, however, that the conclusions in this paragraph, though based on widely held beliefs, cannot be substantiated in a more rigorous way because no adequate empirical research has been undertaken so far.

Although the crisis has not yet been overcome, matters have begun to be gradually sorted out. A constitutional amendment, passed in 1969, made it possible for enterprises to drop managing boards and to experiment with various organizational schemes. Trade unions, authorities and workers have come to realize that certain developments were based on erroneous beliefs concerning various management functions in a labor-managed enterprise. Perhaps the clearest analysis of the mistakes made came from a sociologist, J. Županov (1967a). Županov distinguishes self-management (samoupravljanje), management (upravljanje) and executive work (rukovodjenje). The last mentioned is a partial activity intended to carry out a decision made within a policy framework. The integration of all decisions into a consistent framework is the task of management. But management means only technical coordination, while coordination of various interests, making basic policy decisions, is a task of self-management. Self-management means social integration, the formulation of common goals, which is a precondition for efficient operational work of the management. The confusion between management and self-management generated tendencies to transfer more and more of formal coordination to bodies whose task was social integration. As a consequence, satisfactory social integration was not achieved, while non-professional management meant lower efficiency (Bilandžic, 1969). S. Bolčić has reminded me that this inherently complex problem was complicated even further by a rather naive ideology contained in legislation and political propaganda and advocating direct participation in administrative work as an indispensable of safe-guarding the interests of the workers.

How are the problems encountered to be solved?

Gorupič (1967) and the IES (Institut, 1968b) saw the solution in a fusion of professional competence and self-management. The enterprise may be considered an association of work units. The professional managers of the work units should no longer be appointed, as in the traditional set-up, but be elected by their associates. In this way they would represent the interests of their primary groups, while at the same time being also professionally competent. Managers so elected would make up a managing board which would be both an executive organ of the workers' council and a professional management body. Decisions would be made collectively. Since most of the decisions affecting the daily lives of workers would be made and implemented within economic units and by themselves, executive work would become more and more purely organizational and lose its order-giving character (Novak, 1967, p. 118). Businessmen proved susceptible to this approach (Miletić, 1969). As one might have expected in a country like Yugoslavia, as soon as these ideas had been clearly formulated the practical experimentation began, and the Constitution was promptly amended.

Before closing this section let me note another interesting phenomenon: the development of the so-called autonomous law. Enterprises appear as law-creating bodies. Their self-management organs pass charters and rules governing the organization of work, the composition and responsibility of self-management and other organs, the distribution of income, and the conduct of business. The autonomous law-creating power emanates directly from the Constitution, the rules and regulations are legally binding on all persons to whom they are addressed within an enterprise and disputes are settled by the enterprise organs, except in some specific cases. In this way "a continual narrowing of the area of state law and corresponding broadening of the area of so called autonomous law characterizes the entire process of regulation of social relations in Yugoslavia" (Kovačević, 1969, p. 1).

Enterprise

The introduction of self-management in 1950 implied the dissolution of the centrally planned, administratively run economy. The enterprise was to become independent and autonomous. Individual enterprises needed some guidance and coordination. Therefore so called Higher Business Associations (viša privredna udruženja) were set up in order to replace former state directorates and to preserve continuity in the organization of the economy. The governing councils of the new bodies were composed of representatives of workers' councils of the constituent enterprises. But Higher Business Associations tended to operate along the same administrative lines as former directorates and were therefore dissolved in 1952. A period of *laissez faire* ideology followed. Isolated enterprises were expected to engage in free competition on the market. Attempts to form larger business units and multiplant firms were frowned upon as contrary to genuine self-management and as signs of going back to a disguised state control. In spite of that the system worked well because a special sort of administrative coordination was still effective. The chief coordinator was the Bank implementing the targets of the Plan. The Bank operated a specially designed bookkeeping for every enterprise, distributed the incoming money to various accounts (for wages, taxes, and various enterprise funds) and determined the amount of the necessary working capital which was to be provided on a credit basis, etc. (Vučković, 1952, pp. 11-2p). Although the control was monetary, the value proportions were derived from physical targets.

After 1952 the process of decentralization was not arrested at the level of the enterprise, but went below it. It has already been mentioned that in 1956 the formation of plant workers' councils began and in 1959 the first economic units appeared. The internal cohesion of the enterprise was reduced and it looked as if it was broken up into its component parts. At the same time various monetary and non-monetary administrative controls were gradually being removed. In 1954 the enterprise assumed control of its fixed capital. Fixed assets could be bought and sold without asking for permission. Investment auctions were tried out. In 1958 the enterprise gained control over the internal distribution of income and two years later the trade union control of wages was removed. The stage was set up for a genuine market economy.

As soon as all preconditions for classical free competition of numerous small enterprises were met, it became clear that such an economy would not work very efficiently in the second half of the twentieth century. Since the state refrained more and more from coordinating economic activities, some other agency or agencies had to replace it in that function. That is why the process of integration was initiated. Working collectives themselves had to resume economic coordination in a state that was withering away. The circle of organizational development seemed closed. The process was started by a fully integrated state managed economy, passed through a period of radical decentralization and is now moving towards another stage of full integration in the form of a labor-managed economy.

The forms of integration are various. The simplest one is an agreement for business cooperation intended, for instance, to achieve specialization of the production programs of two or more enterprises. Next comes contractual techno-economic cooperation resulting in joint production, sales or procurement of raw materials. If business relations are numerous and complicated so that it is not possible to regulate everything in advance in a contract, the enterprises form a separate body called a Business Association (poslovno udruženje). By 1962 already one half of manufacturing enterprises were members of Business Associations that first appeared in 1958. In 1967 there were 290 Business Associations consisting on the average of ten enterprises (Dautović, 1968). The next more integrated form is a firm called Affiliated Enterprises (združeno poduzeće). Such a firm is run according to commonly accepted business principles, while constituent enterprises retain operational independence. The latter disappear in a merger. In a seven–year period, starting with 1959 when the process began, the total number of firms was reduced by one half by mergers. It is characteristic, however, that nine–tenths of these mergers were effected within the boundaries of the same or neighboring communes, and only 1.2 percent were interestate mergers. In the same period the number of banks was reduced from 378 to 108. Special status was given to so–called *Unions of Enterprises* (zajednice privrednih organizacija) created for railways, electric power generation and postal and communication services. Membership in these Unions is obligatory. Finally, there are Economic Chambers, organized territorially and associated in the Federal Economic Chamber. The Chambers have a dual role: they help their members in various ways and they also perform a public function, mediating between the state and the business interests. Membership is obligatory.

Mergers and various forms of business cooperation may mean monopoly. That is why a sort of anti–monopoly legislation appeared as well. It is explicitly forbidden to limit free competition in production or sales to any enterprise outside the business group concerned, and government inspectors are expected to take care that there is no sharing of the market or connivance about prices. No serious research about possible monopoly practices has been undertaken as yet, and so there is no possibility of presenting an evaluation here. But it must be borne in mind that the Yugoslav economy will behave differently from other market economies. Workers' management implies a spontaneous public supervision of business conduct and so classical forms of collusion, characteristic of private monopolies, are hardly to be expected, J. Dirlam (U.S. Congress, 1968, p. 3854) finds that the degree of output concentration is higher in Yugoslavia than in the United States; J Drutter (1964) establishes the non–existence of correlation between profits and output concentration and similarly H. Wachtel (1969) finds no correlation between wages and output concentration. In spite of a considerable number of mergers in the period 1959–1963, the degree of concentration actually decreased (Tanić, 1963).

A new enterprise may be founded by an already existing enterprise, by a government agency or by a group of citizens. The founder appoints the director and finances the construction. Once completed, the enterprise is handed over to the work collective which elects management bodies. As long as all obligations are met, neither the founder nor the government have any say about the operations of the enterprise. Enterprises are also to merge or to break in parts. If a work unit wants to leave the mother enterprise, and the central workers' council opposes that, a mixed arbitration board composed of representatives of the enterprise and of the communal authorities is set up. In all these cases it is, of course, implied that mutual financial obligations will be settled.

Since the capital of an enterprise is socially owned, the fundamental obligation of the enterprise is to keep capital intact. If it fails to do so for more than a year, if it runs losses or fails to pay out wages higher than the legal minimum for more than a year, the enterprise is declared bankrupt or the founder undertakes to improve its business record. In the latter case self-management is suspended and replaced by Compulsory Management (prinudna uprava), a form of receivership administered by officials chosen by the commune (Miljević, 1965). Bankruptcies are rather rare because the commune is obliged to find new employment for workers and so prefers to help the enterprise as long as possible.

If integration processes are to proceed efficiently, the organizational forms must be extremely flexible. Thus since 1967 it became legally possible for two or more enterprises to invest in another enterprise and then share in profits. Similar arrangements were adopted in joint ventures with foreign capital (Friedmann and Mates, 1968; Sukijasović and Vujačić, 1968). In an open economy, like the Yugoslav one, foreign capital is welcome provided it does not limit workers' self-management. Therefore direct investment is impossible, but joint ventures are encouraged. The basic motivation for a Yugoslav firm to enter into close business cooperation with a foreign partner is to be found in the desire to secure access to the knowhow and the sales organization of the foreign firm. In this way the Yugoslav firm tries to achieve international standards in technological efficiency and to expand its market.

Theoretical analysis of the behavior of the Yugoslav firm has only begun. Oddly—or understandably—enough, the pioneering work was done by a foreigner, B. Ward of the University of California at Berkeley. In his 1958 paper on the "Illyrian" firm (1958), Ward argues that rational behavior will require maximization of income per worker. In the Marshallian short-run, one product, one factor case this leads to some queer consequences: an increase in wages leaves output and employment unchanged, an increase in fixed costs increases output and employment, and an increase of product price reduces output and employment. In a similar analysis eight years later, Domar showed that by generalizing the production function to include several products and several factors and by introducing the demand curve for labor the results are changed and begin to resemble the traditional conclusions about the behavior of the firm (Domar, 1966). Proceeding along similar lines D. Dubravčić comes to the conclusion that in a labor-managed firm there will be a strong tendency to use capital intensive technology (1967). The empirical evidence does not give unequivocal support to this conclusion. While on the one hand there is a chronic hunger for capital and enterprises use every opportunity to invest, Yugoslav enterprises are also full of redundant workers. Instead of postulating what should be rational, the present author observes the actual practice of Yugoslav enterprises which fix wages in advance for the current year, and at least once a year make corrections (positive or negative) depending on the income earned. If this behavioral rule is used in the analysis, the results are again the same as in the traditional theory of the firm (Dubravčić, 1968). The last in this controversy, Dubravčić, points out that comparative analysis is really not legitimate because it is assumed that a capitalist firm maximizes an absolute magnitude (profit) while a socialist firm is expected to maximize a relative magnitude (income per worker). Dubravčić suggests a symmetrical treatment on the basis of the entrepreneurial input, which is capital in the capitalist case and labor in the socialist case. If a capitalist firm maximizes the rate of profit (profit per unit of capital) it will behave in exactly the same way as Ward's Illyrian firm with entrepreneurial inputs being interchanged. In both cases firms will economize on the entrepreneurial input and this will lead to capital intensive techniques in a socialist firm and to labor intensive techniques in a capitalist firm (Horvat, 1967a) —a nice and almost humorous result.

This brings us to the problem of entrepreneurship in a labor-managed firm. If an entrepreneur is a risk taking and innovating agent—as Knight and Schumpeter would say and most economists would agree—then the work collective qualifies for that role (Horvat, 1964, ch. 6). In fact the work collective is generally treated as an entrepreneur. However, doubts have been voiced as well. Županov argues that the practice of fixing wages in advance means that they are not a residual in the income distribution—as is profit in a capitalist firm—and that this sets up a barrier to entrepreneurial behaviour. He quotes results of empirical research according to which in work units only managers and professionals are prepared to bear risks, while other categories of workers and employees mostly are not. S. Bolčić has drawn my attention to the fact that workers behaved rationally if they were prepared to bear risks only to the extent that they were able to control business operations. That is why managers were both prepared and expected by others to bear risks to a much larger extent. Such as explanation was spelled out explicitly by workers in a case quoted by Lêman (1969, p. 40). In another piece of research undertaken in Zagreb in 1968 it was found that all groups were more prepared to share in losses if output was diminished than if income was reduced while output remained the same or even expanded (Županov, 1967b). On the other hand, it is an empirical fact that wages vary pretty widely depending on the business results. Wachtel quotes data on the issues discussed at workers' councils meetings: two thirds of the agenda items are concerned with general management issues (labor productivity, sales, investment, cooperation with other enterprises, work of management) and only one third with direct worker issues (personal income, vocational training, fringe benefits) (1969, p. 58). Variable wages derived from profits amount to 8–14 percent of standard wages on the average (Wachtel, 1969, p. 100).

The Ownership Controversy

In Marxist sociology ownership relations are the basic determinants of social relations and thus of the socio-economic system. The class that owns—i.e. has an economic control over—the means of production, rules the society. For a long time, and in most instances even today, it has been maintained that private property generates capitalism and state property socialism. In fact the percentage of the national capital owned by the state has been taken as the most reliable measure of the degree of socialism achieved. It follows that a socialist economic policy must be oriented towards an overall economic control by the state and must be hostile towards private initiative.

As already noted, the above described view was generally accepted in Yugoslavia until 1950, and since then it has been thoroughly revised. It is now pointed out that there are at least three reasons why the dogma of the identities between private ownership and capitalism, and state ownership and socialism, is false: the artisans of medieval towns were private owners but not capitalists; in ancient Oriental kingdoms state ownership was frequent and yet that had nothing to do with socialism; in fascist countries the state extensively controlled social and economic life while these countries were obviously capitalist (Horvat, 1969a, Ch. IV). Yugoslav scientists are now quite unanimous in believing that state ownership may be a useful device to initiate socialist reconstruction, but is otherwise as alien to socialism as is private ownership. The present position is well summed up by J. Djordjević (1966, pp. 81, 79): " . . . state ownership of means of production creates a monopoly of economic and political power and . . . makes possible the unification of economic and political power under the control of a social group personifying the state." Thus ". . . the essence of classical (class) ownership is not changed. . . As the holder of the title to property, it (the state) disposes with the producers' labor and its results, on the basis of which surplus labor is appropriated by groups which have their own interests in keeping their commanding functions and thus retaining power and their social status and prestige."

If state ownership fails to promote socialism, what is a feasible alternative? The Yugoslav answer is: social ownership. But the answer to the next question—What precisely is social ownership?—is not so easy and simple. The legal experts agree that social ownership implies self-government, that it is a new social category, that, if it is a legal concept, it does not imply an unlimited right over things characteristic of the classical concept of property, and that it includes property elements of both public and private law (Toroman, 1965, p. 5). In practically everything else there is disagreement. A. Gams and a number of other writers maintain that social property also implies rights of property since property implies appropriation, enterprises are juridical persons and the basic ingredient of the juridical person is property (Gams, 1965, p. 61). Article 8 of the Constitution says that the disposal of means of production in social ownership and other rights over things will be determined by the law. S. Pejović talks about the right of use which is somewhat wider than *usus fructus*, because it makes possible the sale of capital goods, but is narrower than ownership because the right of disposal is not absolute (Pejović, 1966, p. 29). A diametrically opposite view is expressed by Djordjević, and most other writers who maintain that social property represents a negation of property rights (Djordjević, 1966, pp. 84, 90). Djordjević quotes Part II of the Basic Principles of the Constitution to support his view: "Since no one has the right of ownership of the social means of production, no one—neither the socio-political community[6] nor the work organization nor an individual working man—may appropriate on any property-legal ground the product of social labor, or manage and dispose of the social means of production and labor, nor can they arbitrarily determine the conditions of distribution."

Legal writers differ further according to whether they stress the public law or private law component of social property. Further disagreements relate to the subjects of law (state, society as a real community of people, several subjects, no subjects). Next come disagreements on whether social property is a legal, economic or sociological concept or is non-definable in these terms because it relates to quasi-property. And if it is a legal concept, it may be so in various ways. By applying the calculus of combinations we can easily determine the number of possible theories. It seems that available possibilities have been efficiently exploited since M. Toroman (1965) was able to describe thirteen different theories.

The legalistic controversy was somewhat less interesting than the one among economists and sociologists that followed.

---

[6] Territorial political unit such as a commune, a district, an autonomous province, a republic and the federation.

Bajt drew attention to the fact that the legal owner and economic owner may be two different persons. The former holds legal title, the latter derives the actual benefit from the use of a thing (Bajt, 1968). In this sense social ownership implies the non-existence of exploitation which in turn implies the distribution of income according to work performed. If a person or a group of persons are earning non-labor income, they are exploiting others, and in so far as this happens social property is transformed into private property. Thus self-management per se is not a sufficient condition for the existence of social property.

The institution of property already undergoes gradual disintegration under capitalism. Shareholders are legal owners but management exerts real economic control. That is why I prefer to replace the traditional concept of property by a more fundamental concept of economic control (Horvat, 1969a, Ch. 15). The latter always means "control over labor and its products" which is Marx's definition of capital as a social relation (Marx, 1953, p. 167). In this respect legal titles are irrelevant. If artisans or peasants possess no monopoly power—which in an orderly market system is likely to be the case—then they represent no alien elements in a socialist society. And there can be little double that they practice self-management. Horvat and Bajt came to the conclusion that individual initiative is not only compatible with but is an integral part of a socialist system. In fact the process of production can be organized individually or collectively and that is why Bajt talks about two forms of social ownership: individual and collective.

Agreement about the matters mentioned so far is quite universal by now. Differences in views appear when intermediate cases are considered. Yugoslav law makes it possible for artisans and inn keepers to employ 3–5 workers. V. Rašković (1967a, pp. 106–107) and many others consider this to be a form of exploitation, a remnant of the old society, something alien to the system but which has to be tolerated at the present level of development. In support of this view Rašković argues that the employer would not hire workers if this were not profitable for him. It may, however, be argued in reply that a worker, by choosing an individual employer instead of a firm, reveals that he finds such employment more profitable for himself. Such a line of reasoning leads clearly to an impasse. To resolve the question whether workers may be hired by individual employers, and if so how many of them a sociological argument has been advanced as a criterion. As long as an individual employer works himself in the same way as his employees and has not become an entrepreneur merely organizing the work of others, employees may be considered as (often younger) associates in the work process, direct personal relations of a primary group are preserved and the alienation phenomena of wage labor relations are not present.

Discussion of the scope and role of individual work was invited by political bodies and very soon decisions were made following more or less the ideas expounded above. Individually organized production became a constituent part of a socialist economy.[7]

IV. *Market and Prices*

Price Policy

Price Policy represents an incessant series of attempts to control the famous law of value (supply and demand relations). Its history is instructive since it provides an insight into the working of various institutional arrangements.

---

[7] The private sector—which Yugoslav economists prefer to call "individual sector" in order to avoid various connotations of the attribute "private"—accounts for 29 percent of GNP and this percentage has not changed in the last fifteen years.

Administratively Set Prices: Immediately after the war, with the economy almost totally destroyed, there was an extreme scarcity of all goods. The prime purpose of economic policy was to prevent profiteering and to generate output by any means available. This was the period of "profitability at all costs" (Radulović, 1968, p. 143).

Prices were determined on the free market only for a few luxury products. Mostly prices were set on the basis of actual costs incurred and could vary from one producer to the other. The Price Offices would examine each case and make the relevant decisions (normiranje cijena). This was not a very efficient procedure. Since actual cost was taken as given, there was no incentive to economize on inputs. Wages were fixed, and products could always be sold. In order to minimize risk, producers tended to inflate costs in their price proposals and in order to keep prices down the Price Offices tended to apply linear reductions to proposed prices. The authorities and the businessmen began to play at hide-and-seek, which is so characteristic for an administratively controlled economy.

The launching of the First Five-Year Plan in 1947 required a system of uniform prices (jedinstvene cijene). Uniform prices were determined by the planning authorities and were expected to be rigidly stable. The aim was to provide a link between the physical and the value part of the plan, to have a control over the implementation of plans and to avoid the administrative costs of changing prices frequently. Prices were formed by adding an average rate of profit to average cost for a product. The less efficient producers had planned losses, the more efficient ones extra-profits; in both cases differences were settled with the budget. Through the establishment of the system of uniform prices, the law of value was considered to be subject to an efficient social control (Kidrič, 1948, p. 143).

It soon became evident that uniform prices did not equilibrate supply and demand. There was chronic excess demand. Private producers (peasants and artisans) held a large share of the market and their incomes could not be easily controlled. Most consumer goods were rationed and sold at the existing uniform prices, but available quantities of consumer goods were not sufficient to satisfy the needs of the entire population at the lower uniform prices. By the end of 1947 the first quantities of consumer goods were supplied to the free market at higher uniform prices (više jedinstvene cijene). These prices were derived from the existing uniform prices by applying multiplying factors varying from 2 (for potato and beans) to 6½ (for garments). The resulting trading profit was absorbed by the budget. In 1948 about 45 percent of consumer goods were supplied at higher uniform prices (Šefer, 1956, p. 376). In this way, it was hoped, excess money incomes would be absorbed.

In agriculture a system of compulsory deliveries (obavezni otkup) was applied. Peasants were obliged to sell most of their products to the state at prescribed low prices. For the money they obtained they could not buy all those industrial products they wanted. Thus they tried to reduce deliveries and substitute their own consumption for money incomes. The government reacted by creating a market for industrial goods at higher uniform prices. Peasants reciprocated by evading compulsory deliveries and supplying more goods to the free peasant market, the only section of the market where the prices were equilibrating supply and demand. These prices tended to rise fast and so the government decided to substitute a carrot for the stick: in 1948 the government introduced linked prices (vezane cijene). Agricultural prices were linked with industrial prices in such a way as to establish the pre–war parity. Peasants sold their products to the government at lower prices and in return obtained coupons which enabled them to buy industrial products at prices that were about 16 percent lower than commercial prices (Dobrinčić *et al.*, 1951, p. 141).

Local markets were less rigidly controlled. After 1949 local enterprises could in principle sell their products at commercial (higher uniform) prices. Trading establishments that were supplied by two different producers—national and local—were now unable to sell commodities at one single price. And so sliding prices (klizave cene) were invented. The selling price slides in a span determined by the lowest and the highest supply price. These prices were approved by the local authorities. Thus two different markets were created: one for enterprises that traded at lower and higher uniform prices, and the other for retail trade and population where prices approached free market prices.

The system of linked prices did not work too well. The supply of industrial goods was inadequte and richer peasants began to speculate with coupons. In 1950 only some agricultural products could be sold at linked prices. More of the peasants' products went to the free peasant market whose counterpart in the state sector was the system of sliding prices. Higher uniform prices, being administratively set, were lagging behind the free market prices. Output of consumer goods was stagnating, even falling, while incomes were rising (Čobeljić, Mihajloić and Djurović, 1954, p. 49): (see table 4A).

The widening gap between supply and demand could be controlled by administrative or economic means. The government chose the latter. In the transitional year of 1951 there were eight different price categories coexisting simultaneously (Dobrinčić *et al.*, 1951, p. 143). Sliding prices were superseded by higher prices for consumer goods. Rationing was abolished. Consumer goods prices were left to be regulated by the market while producer goods prices were increased one to twelve times and then frozen for about half a year. In 1952 compulsory deliveries of agricultural products were abolished. By the second half of 1952 all prices were freely formed with the exception of a few goods (bread, sugar, electric power etc.) for which ceiling prices were established.

Development of the Market: The strategy of the 1951/1952 price reform can be summarized as follows: (*a*) a sufficiently large increase of prices to absorb all excess money incomes; (*b*) a sufficiently large increase of retail prices of manufactured consumer goods relative to agricultural prices to generate the capital accumulation necessary for fast growth, (*c*) a smaller increase in producer goods prices to stimulate investment and the expansion of compartment I (producer goods industries). The first two goals were achieved with remarkable success. As a result, industrial producer prices were kept stable over a period as long as a decade. The third strategy proved to be deficient and generated a lot of trouble.

While the general index of industrial producer prices was declining for almost three years, prices of certain raw materials (ferrous and nonferrous metallurgy, building materials, wood products were rising.

TABLE 4A

	1948	1949	1950	1951	1952
Consumer purchasing potential	100	128	125	245	327
Retail trade in real terms excluding peasant trade	100	100	94	70	77

That is why in 1954 ceiling prices were set by the government for a number of raw materials, and in the next year the list of controlled intermediate goods was further extended. In 1955 industrial producers prices rose by five percent, which led to the creation of the Federal Price Office in the same year. Since then a system of administrative control of prices has been gradually developed. The essential features of this control are as follows:

1. The government sets *fixed* prices for electrical power, cigarettes, transportation rates, sugar, oil, salt and some other commodities.
2. The government sets *ceiling* prices for metallurgical products, coal, petroleum and some other goods.

These two categories of prices are changed at infrequent intervals. But when they are changed, the change is rather drastic.

3. Control on the basis of *prior price registration* is the most frequent kind of control. It was introduced in 1958. Producers intending to raise prices are obliged to notify the Federal Price Bureau thirty days beforehand. If within this period the FPB does not veto the price increase, it can be effected. The principal criteria for placing a product under control are: (a) its importance for the standard of living or for production costs of other products; (b) scarcity on the market and (c) the monopoly position of the producer (Vuković, 1968).
4. *Control of trade margins* is implemented by republics for wholesalers and by local authorities for retailers.
5. *Price freeze.* This instrument was used only on two occasions, in 1952 and in 1965, during two price reforms.
6. Agricultural prices are placed under a special regime. *Guaranteed prices* are applied to staple food products. This means that the Federal Food Reserve Board is obliged to purchase all quantities of the products offered for sale and to pay guaranteed prices. For milk and industrial crops *minimum* prices apply. This means that if these products are bought at least the minimum prices have to be paid for them. An industrial crop is normally not grown unless the producer has a prior contract with the buyer. Prices used in such cases are *agreed–upon prices.*

Industrial prices have been most heavily controlled. In the last decade this control was exercised over the following percentages of the value of industrial output (Radulović, 1968, p. 282; Drutter, 1968, p. 113; Institut, 1969, p. 6):

TABLE 5

1958	31.2	1965	70
1962	67.0	1966	66
1962–65	60.0	1967	53
		1968	46

The time series of prices, given in Table 6 may give an idea of how efficient the price policy and price control were.

After 1961 the administrative control of prices was increased and so was the inflationary pressure. What in fact happened?

The most frequent form of price control—prior price registration—could not be adequately applied to new products. By making small changes in the design of a product an enterprise would transform it into a new product and so could evade price control. In 1964 almost twenty five thousand new products were launched. Low and rigidly controlled prices of raw materials made their production unprofitable and so depressed output; in agriculture prices were particularly depressed. That is why in 1964 prices were raised administratively in agriculture, the food processing industry, energy generation and nonferrous metallurgy. Next, differential taxation, a system of premiums and subsidies, and administrative interventions in foreign trade tended to preserve and even increase price disparities (Pertot, 1966). As

TABLE 6.—CHANGES IN PRICE LEVELS IN PERCENTAGES PER YEAR

	1952–1963	1964	1965	1966	1967	1968
Producer prices in manufacturing and mining	+0.9	+5	+15	+11	+2	0
Agricultural producer prices	+8.6	+24	+43	+16	−3	−4
Retail prices (including services)	+3.9	+9	+29	+23	+7	+4

Sources: *Jugoslavija 1945–1964. SGS–1969.*

a consequence individual enterprises conducted their business under highly unequal conditions. Producers whose prices or wages were lagging behind were trying to catch up with their neighbors. The Federal Price Bureau received 12,800 requests for price increases in 1961 and 69,000 requests in 1964 (Drutter, 1968, p. 107). But the most important reason for the break in price trends in 1961 lies elsewhere. Until 1961 personal incomes were quite efficiently controlled by fiscal and nonfiscal means (Trade Unions). That is why prices were quite stable (except in agriculture) and administrative controls relatively few (Institut, 1968b, pp. 37–41). In 1961 income controls were abolished, very soon a cost–push inflation occurred and, despite increasing administrative control, prices went up. A few years later the Institute of Economic Studies suggested that the Federal Price Bureau relax administrative price control and focus its attention on income control (Institut, 1969, p. 41). The suggestion was not followed and instead monetary policy was used as the chief anti-inflationary weapon.

By 1965 the economy was ripe for another radical price reform. In March prices were frozen and a tax reform carried out. Various subsidies were drastically reduced and the tax burden of enterprises alleviated. In the next few months a new price structure was prepared. In July the dinar was devalued; new prices were introduced and frozen. Relative prices of certain raw materials, intermediate goods (electric power, petroleum, ferrous and nonferrous ores and metals, chemicals, timber products and agricultural products) and transportation services were substantially increased. World prices (as registered in exports or imports) were taken as a basis for the new price structure. This was to make possible a rapid integration of the Yugoslav economy into the world economy. World prices were corrected upwards or downwards by taking into account capital accumulation needs of various industries and other specific purposes. A new customs tariff was to iron out these differences.

The price stabilization proceeded rather slowly, as can be seen from Table 6. The lifting of price controls went even slower. In 1968 prices looked stabilized, but almost one half of industrial prices were still under control. Disparities between controlled and uncontrolled prices began to emerge. The output of certain industries tended to become depressed. In 1969 prices began to rise again. The experience of 1964 seems to have been repeated. The reform of 1965 eliminated the worst price disparities, but subsequent price controls created new ones. The price game seems far from being successfully completed.

There has been a lively discussion about the appropriate pricing system for a labor managed economy. This discussion hardly touched the classical controversy on marginal cost versus full cost pricing. Since marginal cost pricing requires government intervention, the lack of interest in this procedure among Yugoslav economists is understandable. On a more theoretical level it was pointed out that allocational efficiency—as represented by marginal cost pricing—is inferior to growth efficiency—as represented by full cost pricing which makes possible the business autonomy of an enterprise (Horvat, 1964, Ch. 2).

The price debate was centered around the problem of how the price is to be formed. It started in 1950 when Kidrič opted for the "value price" (1950b). In his last writing in 1952 the late Kidrič described the value price as the one consisting of costs of production (including wages) and accumulation (gross profits) calculated as proportional to wages. These prices were actually tried out in 1953 and 1954. Kidrič compared the rate of accumulation principle with the traditional average rate of profit principle (profit proportional to capital invested is characteristic of Marx's price of production) and came to the conclusion that only the former was appropriate for a labor managed economy. In his view the average rate of profit principle "represents a contradiction to socialist planned management of the economy," and leads to "a kind of cooperative capitalism" (Kidrič, 1952 pp. 42, 46). A decade later M. Todorović—who was to become the secretary of the League of Communists—came to the opposite conclusion. He maintained that in a system of commodity production, including its socialist variety, in which fixed capital is used, prices must take the form of prices of production. Since capital is socially owned and production is planned, the use of prices of production cannot lead to the same consequences as in a *laissez faire* framework of liberal capitalism. (Todorović, 1965, pp. 60, 65, 78).

Strange as it may sound, there is no basic disagreement between Kidrič and Todorović. The difference between their views primarily reflects the difference in the degree of economic sophistication. In 1952 Kidrič's view was commonly accepted—by Todorović as well—while today hardly anybody would be prepared to support it. Todorović's theory of the specific price of production (specific because social planning is one of its basic ingredients) as an equilibrium price in the Yugoslav setting has been accepted by a certain number of economists—Z. Pjanić, V. Rakić, Maksimović (Institut, 1968)—but by no means by all. In a heated debate in Sarajevo in 1964 another group of economists—Korač, Sirotkovic, Dabčević, T. Vlaškalić—expounded the theory of "income price" (Savjetovanje, 1964). In their view the Yugoslav enterprise maximizes income in relation to suitably defined inputs. Other economists were busy inventing new types of prices: gravitational (Mesarić, 1965), normal, actual, social reproduction price (Černe, 1966, p. 233), etc. Radulović was able to describe six different price theories of this sort (1968, pp. 299–326).

Price theory is closely linked to distribution theory which we shall consider in the next chapter.

Distribution Policy

It is not conventional to talk about distribution policy. One is accustomed to speak about wages *policy* and distribution *theory*. However, as we proceed, it will become evident that in the Yugoslav setting distribution policy is also a meaningful concept.

Wages Policy: In the administrative period 1945–1952 workers were government employees classified in a certain number of salary categories according to their skills. Directorates set work norms whose overfulfillment brought an increase in pay. Managerial personnel would get premiums for the fulfillment of the government plan. The salary span was 1:3.5 (Tomic, 1968, p. 6), as compared with 1:16 before the war (Bilandžić, 1967, p. 56).

The lack of material incentives was compensated for by moral incentives such as public praise, the trophy-flag, the title of shock worker or of innovator. In the post-revolutionary atmosphere these incentives were very powerful.

After several years the lack of material incentives became a serious obstacle to efficient production. Due to post-war scarcities and to an egalitarian ideology, by 1953 the salaries of office employees in industry had been reduced by one third and of civil servants by one half relative to workers' wages and compared with the prewar levels (Berković, 1969, p. 81). Non-wage income "fringe" benefits) was higher than wage income. Since 1952 both trends have been reversed. Trade Unions advocated higher skill differentials. Economists (Bajt, 1956) urged an increase in the share of discretionary income (income after taxes and contributions left to free disposal of an enterprise) in order to increase productivity. Wachtel finds that interskill differentials increased until 1961 and then began to fall. The average income span between the highest and the lowest paid job is now 1:4 (Berković, 1969, p. 82). M. Janković estimates that wage income increased to 65 percent of total workers' income in 1956 and to 73 percent of total income in 1967 (1968, p. 159). The idea was to leave to the market the job of determining the appropriate income differentials and to stimulate efficiency by increasing the discretionary part of workers' income. The latter was also thought necessary in order to curb centralist distribution of income.

Since 1952 it has been the task of workers' councils to determine wage differentials and work incentives. The distribution of income between the enterprise and the community was settled in a very simple way. On the basis of the social plan, the expected income of the enterprise and the corresponding wage bill were determined. The difference between gross income (depreciation excluded) and wages was called accumulation and funds (AF). The ratio between AF and wages was called the rate of accumulation and funds. This rate was applied to actual gross income earned in order to derive wages. It was mentioned in section 10 that the AF rate was considered an appropriate socialist substitute for the rate of profit, and that was its theoretical justification; whatever the merit of that argument, the practical effects were good. The AF rate helped to bridge the institutional gap between complete administrative control and a relative autonomy of the enterprise. It also induced workers to economize on labor. In 1953 employment in manufacturing and mining increased by 5 percent, and labor productivity by 6.2 percent. In 1954, when the AF system was abandoned, employment increased by 13 percent and labor productivity slightly fell.

The AF rates were, of course, not uniform. The 1952 plan envisaged a rate of 19 for agriculture and another of 582 for manufacturing and mining. This difference reflected the already described goals of price policy: industrial prices were inflated in order to facilitate the collection of investment resources. However, even within manufacturing different industries had widely different rates. In industries with high rates there was no incentive to reduce costs. Since the rates could not be established very precisely, some collectives began to earn high wages. The government reacted by introducing a tax on "the surplus wage fund" (the difference between the standardized and the achieved wage bill). Since the standardized wage bill was the product of an average wage rate and the number of employed, the enterprise increased employment—often fictitiously—of less skilled workers in order to reduce the tax basis. The government reacted by differentiating taxation according to skill categories. Enterprises countered by artificially changing the skill structure, declaring their workers to have higher skill.

The AF rates were clearly a not very refined instrument of economic policy. They were introduced in the belief that they could be standardized for all enterprises within an industry group. Soon, however, individual rates had to be prescribed for each particular enterprise. This implied direct administrative interventions which were at variance with the basic intentions of the new system. In 1954 the AF system was replaced by a system called "accounting wages," which lasted for the next three years.

Yugoslav economists had been complaining for some time that in their economic calculus enterprises do not consider capital services as a cost item (Lipovec, 1954, p. 142). That was a natural result of the fact that capital was given to enterprises free of charge. This practice was discontinued in 1954 when a capital tax of 6 percent was introduced. This tax was considered as a price for the socially owned capital and was also levied on capital invested from enterprise funds. Apart from that, the enterprise was obliged to pay normal interest rates on credits granted by the bank. Also the profit and turnover taxes were introduced, the latter becoming the chief instrument of accumulation. In this way instruments of economic policy became more varied and more flexible.

The new system implied a division of the wage fund into two components: accounting wages and wages out of profit. Accounting wages were derived by applying prescribed wage rates to skill categories taking into account actual working time. Again skills were fictitiously increased. Working time as a basis of accounting led to a disregard of work norms. The next year "wage schedules" (tarifni pravilnik) were introduced. Wage rates were determined by the social plan. Wage schedules of individual enterprises represented a kind of collective agreement between the enterprise and the Trade Union and local government (Tomić, 1968, p. 11). Differential efficiency was accounted for, and a part of profit was used as a premium for improvements of quality, reduction of costs etc. Since profit was taxed at 50 percent, enterprises tried to reduce profit by increasing wage rates and reducing norms. The government commissions for wages were unable to prevent this from happening.

In 1957 the First Congress of Workers' Councils was held. The Congress asked that the autonomy of the enterprise be widened. This primarily implied greater independence in income distribution. The division of income into wages and profit was considered inappropriate and reminiscent of wage–labor relations. In order to meet these demands, in 1958 the income distribution system was changed and a compromise reached. The wage schedules remained and were still subject to approval of local authorities and trade unions. The enterprise income was treated as one single whole and was distributed by workers' councils into wages and contributions to various funds. The difference between income and accounting wages (called minimum personal income) was progressively taxed. The wages in excess of the basic pay were also progressively taxed (Pejovich, 1966, pp. 98–99).

Progressive taxation was very much resented. And so was the outside tutorship as far as wage differentials were concerned. In 1961 both were abolished. Workers' councils became completely independent in determining wage rates and distributing income. Progressive taxation was replaced by a flat 15 percent levied on income. In 1965 even this tax was abolished.

Changes in wages policy implied drastic changes in relative factor shares. If we divide value added into gross wages (wages and taxes levied on wages) and gross rentals (depreciation, interest, net profit and taxes levied on capital), the percentage share of the latter in manufacturing and mining varied as follows (Horvat, 1969b, p. 41):

TABLE 7

1952	10%	1961	54%
1953	11%	1963	53%
1955	74%	1964	50%
1957	77%	1965	48%
1959	67%	1966	46%
1960	62%	1967	45%

Percentage shares of gross wages represent, of course, complements to 100% of the figures quoted for rentals. In the AF system depreciation was the only capital cost. The introduction of profit and capital tax in the system of accounting wages increasedcapital cost drastically. The gradual reduction and final elimination of profit taxes, which implied a relative increase in wage tax, reduced the share of gross rental to somewhat more than one half of the value added. On these changes price changes were superimposed. The increase of food and services prices after 1960 increased nominal wages; the abolition of various subsidies at the same time and in particular after 1964 made possible a reduction in taxation which to a certain extent offset the effect of wage increases. The next effect was to lower the share of gross rental below 50 percent. The adaptation of an enterprise to these changes required an extraordinary effort on the part of the management. But enterprises did react. Simultaneously with increased capital charges the capital coefficient (the ratio of gross fixed capital to gross material product) in manufacturing and mining fell from 3.6 in 1955 to 2.5 in 1964 (Horvat, 1969b, p. 51). If enterprises are market oriented and if the production function is linear homogeneous (which proved to be an acceptable approximation), the elasticity of output with respect to capital in the last decade must lie somewhere in the region 0.45 to 0.62. The actual elasticity coefficient turns out to be 0.48. This is taken as one indication that the economy is following market rules (Horvat, 1969b, p. 42).

While wage systems with wage schedules and progressive taxation were applied, real wages lagged behind productivity increases and producer prices were stable. From 1958 on real wages began to increase faster than labor productivity, and the discrepancy between the two series was widened particularly in the cyclical trough in 1961/62 and after 1964 (Popov, 1968, p. 627). The peculiar movements of prices that followed were considered in the section on Price Policy. Another peculiarity was established by Wachtel: interindustry wage differentials continued to increase, and interindustry wage structure appeared as a function of average productivity which explained 80 percent of the variance (Wachtel, 1969, pp. 151, 175). Popov found a high correlation between the rate of growth of industrial output and the productivity of labor ($r = 0.86$) (Popov, 1968, p. 622). If all these bits of information are put together, the following interpretation begins to emerge.

Trade Unions announced the principle: wages should increase proportionally to the productivity. The principle was widely accepted, and it is a sound principle when applied to the economy as a whole. If applied to individual enterprises, it generates great trouble. In a rapidly growing economy various industries expand at widely different rates (petroleum industry at 19.2 percent, tobacco industry at 5.1 percent per annum in the period 1952–1966). Thus rates of growth of labor productivity are bound to differ very much (11.7 percent and 1.2 percent respectively). Thus wages must differ and differentials must increase in time (money wage rates increased 12.8 times in the petroleum industry and 8.3 times in the tobacco industry in 1952–1966) (Popov, 1968, p. 630). Kovač found that in 1966 wage rates for the same category of skill in the highest paid and the lowest paid industry group were related as 2:1 (1968, pp. 130–33). All this is, of course, in flagrant contradiction to the principle of distribution according to work. That is why Bajt remarked that the principle of remuneration according to productivity actually denied the principle of remuneration according to work performed (1967b, p. 363). Deviations of productivity income from labor income have been analyzed by the present author. They represent (after deductions for other facto costs) a form of rent which I call the rent of technological progress (Horvat, 1962b). The faster the rate of growth, the more important this rent becomes.

Rašković (1967b, p. 230) and others suggested that the principle of distribution according to work be replaced by a more appropriate principle "according to the results of work." It is not the process of work as such but its results that have to be rewarded. Rašković noted that grossly imperfect markets in Yugoslavia meant exploitation of one group of collectives by another, more privileged, group (1967b, p. 218).

The meaning of the principle, "according to the results of work" has been stretched by Šefer in a rather curious fashion. Šefer notes that in developed capitalist countries free market wage determination has been replaced more and more by a policy of "equal pay for equal work." He feels that such a policy is inapplicable in Yugoslavia because workers bear business risks, i.e., they share in both profits and losses. Work cannot be remunerated automatically; it has to be socially recognized, which happens at the market where the exchange determines the result of work. The principle "equal pay for equal work" could be implemented only in a system of state ownership and state management of the economy (Šefer, 1968b, pp. 74–75). Thus Šefer, Korač and a certain number of others in fact argue that the principle, considered Marxian, can be implemented in a capitalist and étatist setting, but not in a self-government system. The fallacies of this *laissez faire* reasoning are obvious: market imperfection provides no criteria for the social recognition of somebody's work; the redistributive effects of market imperfections can be eliminated also by means other than the étatiste ones.

Other Issues: On income differentials due to technological and other rents, differentials due to variable entrepreneurial abilities of various working collectives are superimposed. Šefer quotes data for Belgrade enterprises in 1967 when the same jobs in various enterprises were paid rates as different as 1:3 or 1:4 (1968a, p. 434). It is clear that such extreme differences generate enormous inflationary pressure. There is also an additional consequence. Capital intensive enterprises are able to improve their personal income position by distributing a part of profit in wages. That is why wage rates are positively correlated with capital intensity. Yet, if profits tend to be reduced, enterprises become more and more dependent on outside sources for financing their investment. This generates new difficulties which we will consider in the section on Banking and Monetary Policy.

Apart from technological rent, the classical forms of rent were both discussed in the literature and applied in practice (Bakarić, 1950; Horvat, 1953). Agricultural rent is absorbed, in principle, through taxation according to cadastral revenue. Mining rent represented a separate item of income of mines and crude oil producers for several years. However, it was determined in a rather arbitrary way and generated regional differences. Consequently, it was resented by the enterprises and was eventually abolished. All urban land belongs to communes and urban rent is used to finance communal investment.

J. Dirlam, an American student of Yugoslav economic affairs, points out that the Yugoslav system can be viewed as one in which labor employs capital, instead of a system in which capital employs labor as is the case under capitalism. The social ownership of capital requires a somewhat different approach to capital charges in the labor-managed enterprise as compared with its capitalist counterpart. The floor and not the ceiling is set for depreciation rates. Profits need not be taxed and instead payroll taxes are suggested (Institut, 1968b). A tax on capital is primarily an instrument for allocating resources and not necessarily a device for collecting revenue for the government. The revenue from capital taxation has been used by the government to finance major investment projects and also to finance the Fund for underdeveloped regions. Resentment against these redistributive activities of the Federal Government has been growing, and recently a political decision has been made to abolish the capital tax. Many economists disagree with this decision. Some argue that the abolition of capital tax, which represents the price for the use of social capital, will initiate a transformation of social ownership into collective ownership. D. Gorupić and J. Perišin argue that the price of a product should contain an element of growth (1965, p. 124). This is to be achieved if accumulation is determined by the social plan in the form of interest on capital used. But this money must not be expropriated by the state; it ought to remain in the enterprise earmarked for investment. Thus this internal interest is to be treated in the same way as depreciation. In order to cope with business fluctuations, minimal depreciation cum accumulation must be determined in a cumulative fashion (Gorupić, 1968, pp. 12, 13). Lavrač maintains that the accumulation-protecting interest rate may be differentiated according to industries and regions (1968). S. Popović suggests that the compensation for the use of social capital will provide the bulk of development resources. After all factors of production, except labor, are paid their shares, the remaining net income is to be distributed among workers. Additional accumulation can be derived only from this private income, which means that workers remain owners of that part of capital (S. Popović, 1968). Similar is the position of Černe who maintains that the participation of workers with their own means in the development of the enterprise—which implies receiving adequate interest or dividends—would stimulate rational behavior of workers and management bodies (1967a, p. 21). On the other hand, Samardžija argues that this is both economically irrelevant and socially dangerous. Contemporary shareholders participate in the profits of their corporations with only small percentages that accrue to dividends. And attempts to make workers co-owners must end in the establishment of a separate group of owners of means of production within the society (Samardžija, 1968, pp. 145, 303).

We have thus reached the point at which the general principles of an adequate distribution policy may be discussed. There seems to be considerable agreement on two issues. (1) As great a part of income generated as possible should remain under the direct control of the working collective. (2) Only labor income should be distributed in wages. These two principles imply a sharp division of income into two components: labor income appropriated by individual workers and nonlabor income belonging to the society but remaining under the control of the working collective and used exclusively for investment purposes.

In order to be able to divide net income into its labor and nonlabor parts, we need a theory of factors of production. In this respect Bajt follows the traditional approach and defines factors of production as sources of productive services. He enumerates five such sources: labor, entrepreneurship, invention, land and capital (1967b, p. 351). The first three generate labor income, although normally a small proportion of income from inventions is appropriated by inventors. This theory leads Bajt into difficulties when he has to explain monopoly income. He then argues that in a market economy monopoly participates in income; monopoly does not add to output but only adds to income of all factors (1967b, p. 357).

In order to avoid the shortcomings of the traditional theory, the present author defines factors of production as types of forces that influence the generation of output. Factors have to be priced in such a way as to lead to an optimal allocation of resources. The latter means achieving maximum output from given resources or minimum input of resources for a given output. There are four factors: labor, entrepreneurship, capital and monopoly. The first two generate labor income (wages and profit), the latter two generate non–labor income (interest and rent). Creative work and organizational work as well as routine work generate labor income. The income due to the activities of the work collective as a whole represents entrepreneurial income. Capital services are priced in the usual way and have already been discussed. A few more words need to be said about the morphology of rent. Rent is the price of monopoly in the sense that it represents the surplus over the minimum supply price of resources. Land rent appears in three forms described by Marx (differential rent I and II and absolute rent), then there is mining rent and a somewhat special urban rent. The rent of technological progress—due to the fact that certain industries expand faster and enjoy economies of scale effects, or participate more in general technological advance, or both—has already been described. Bajt adds the rent from market monopoly, which he describes as a situation when the selling prices are above normal and the buying prices are below normal (Bajt, 1962, p. 93). After land, natural resources, technology and market monopolies are accounted for, the remaining part is a monopoly in the narrower sense. Except for the last, the prices of the other monopoly factors may be in principle determined either by the market mechanism (land and mines) or by economic analysis (technology and market). As far as the latter is concerned, progressive taxation may in practice prove a more efficient procedure. If taxes are designed in such a way as to be generally considered as just, they will not affect the supply of resources and this is how in fact we defined rent (Horvat, 1964, ch. 3, 4, 6).

Actual business practice and legislative measures do not quite follow the principles discussed above. The productivity–wage practice leads to an appropriation of a considerable part of non–labor income. The same consequences follow from the facts that mining rent is included in undifferentiated income and that there is no progressive taxation. In 1968 the new law on the distribution of income in the enterprises included income from capital invested in other enterprises in the undifferentiated income of the collective–investor. P. Jurković promptly called that a rather dubious theoretical solution (1969, p. 50). In general, the distribution of income according to the work performed is still a goal to be reached.

Foreign Trade Policy

Background: The pre–war trade structure was rather simple. Food and other agricultural products represented about one half of total Yugoslav exports. One fifth of exports consisted of wood and almost an additional fifth of non-ferrous ores and metals (Dobrinčić *et al.*, 1951, p. 408; Fabinc *et al*; 1968a, p. 144). Thus close to ninety percent of export earnings were provided by these three sectors producing raw materials and semi-manufactured goods. Immediately after the war the development strategy consisted in (1) expanding the exploitation of natural resources in these three sectors and (2) in using the export proceeds to finance imports of equipment and other producer goods. It was also expected that (3) the Soviet Union would provide great help in speeding up economic development. The second part of the program was carried out successfully, the share of consumer goods in imports was reduced from 22 percent before the war to only 11 percent in the period 1947–1951 (Čehovin, 1960, p. 59). The first and the third parts encountered unexpected difficulties.

Due to a decline of per capita agricultural production and rapid industrialization, agricultural export surpluses were reduced and so was the total volume of exports. It soon became fashionable to explore the question whether Yugoslavia was not becoming a permanent net importer of agricultural products (Srdar, 1953). The nationalization of foreign property imposed a new burden on the balance of payments. On the other hand, reparations for war damages were to a large extent left unpaid, in particular by Western Germany and Hungary. Immediately after the war about 75 percent of foreign trade was conducted with the Soviet Union and her East European allies. In 1947–1948 the trade shares with these countries were stabilized around 50 percent in exports and 42 percent in imports. In the middle of 1948 the ominous Resolution of the Cominform meant the end of good relations. By 1949 the Soviet group reduced the trade to one-third and in 1950 it was cancelled altogether. The Soviet Union and her allies applied a total boycott to all relations with Yugoslavia.

Thus the country was cut off from the East completely. It was separated from the West as well, as it did not enjoy the facilities mutually provided by western countries to each other. It was not included in the Marshall Plan; it remained outside GATT. In short it was isolated in a hostile world. The five-year industrialization plan—imbued with so many hopes—had only been initiated, when suddenly the contracts were broken, and supplies of equipment and materials ceased to arrive. Trade was declining:

TABLE 8

	1948	1949	1950	1951	1952	1953	1954	1955	1956	1965
Exports	100	79	74	64	87	80	102	99	122	328
Imports	100	95	86	114	115	106	103	130	142	311

Sources: *Jugoslavija 1945–1964*, p. 77: *SGJ*–1959, p. 121.

Foreign exchange reserves dropped from 43 percent of the value of imports in 1937 to 12 percent in 1948 and to 4 percent in 1952 (Mrkušic, 1963, p. 186). Personal consumption was declining. Defense expenditures amounted to twenty percent of national income. Two severe droughts, on in 1950 and the other in 1952, proved unexpected allies of the Cominform and reduced agricultural output to 25 percent below the pre-war average. The situation looked hopeless. That is why Stalin expected surrender.

Yet this nation was not accustomed to surrender; it was more at home in fighting back. And it did so, for the first two years struggling practically alone. Investment plans were changed, trade was channelled towards the West, even the economic system was changed. From 1951 on, foreign economic aid began to flow, mostly from the United States. It consisted primarily of food, raw materials and military supplies. The aid amounted to 38 percent of total imports in 1951, and over the next decade was gradually reduced to zero.

The crisis was soon overcome and the economy entered a period of unprecedented growth. The effects of the heavy capital investment of the First Five-Year Plan began to materialize in rapid expansion of industrial output. The new agricultural policy soon generated phenomenal growth of agricultural output. Exports were catching up with imports. In 1954 the first trade contacts were established with the East European countries. After the conciliatory visit of Premier Khruschev to Belgrade in 1955, normal trade relations were established and so a precious outlet for increasing exports was found (Obradović, 1962, p. 40). In the decade that followed, exports increased 3.3 times, i.e., at a rate twice as high as in the world as a whole.

These developments were too good to last long. In 1957 the Common Market was born in Rome. Two years later EFTA was created in Stockholm. Practically all West European countries became members of the one or the other trading group. East European countries belonged to COMECON, created in 1949, but actually operating since 1954. Yugoslavia found herself isolated again. At first it did not matter too much. But gradually intrazonal trade in all three areas began to increase rapidly and to depress trade with third parties. This was true in particular for the Common Market, the most important trading partner of Yugoslavia. Common Market countries account for 30 percent of Yugoslav exports, 38 percent of imports and two-thirds of financial transactions. What makes this trade so vulnerable is the fact that between one third and half of Yugoslav exports to Common Market countries consists of agricultural products. Regular and variable import tariffs in the Common Market amount on the average to 50 percent of the Yugoslav export prices, for beef even to 60–70 percent, which clearly cannot encourage exports. Variable protection rates, when first announced to GATT, were said to be an exceptional instrument, the customs tariff remaining the basic one. In fact, however, variable rates amount to 2.5 times the regular tariff, they are changed daily, weekly or quarterly and represent a permanent instrument of total protection (Žiberna, 1969; Mitić, 1969).

Yugoslavia reacted to the new situation by trying to increase her trade with the developing countries. This attempt met with a limited success. Imports from developing countries increased to a maximum of 14.1 percent of Yugoslav imports in 1964 and there has been a permanent balance of payments surplus with these countries (Pelicon, 1968). Next, close relations were established with GATT. At first an observer, Yugoslavia became an associated member of GATT in 1959 when she also enacted the Customs Law. In 1961 a temporary customs tariff was produced and next year Yugoslavia became a temporary member of GATT. In 1965 a new, permanent customs tariff was enacted, and a year later full membership was granted by GATT.

COMECON was also approached. Its members absorb almost one third of Yugoslav trade. In 1964 Yugoslavia became an observer in COMECON. With the Common Market special agreements are negotiated.

India and the United Arab Republic account for one third of Yugoslav trade with developing countries. In 1966 the heads of the three countries initiated a scheme which became known as Tripartite Co-operation. The agreement, ratified in 1968, comprised 500 products to which preferential rates of 50 percent became applicable, and envisaged also industrial co-operation. It was also suggested—this time by economists and not be politicians (Bilandžić, 1967, p. 33)—that a Danubian trading area be formed. If that had proved possible, it was hoped that the area could have been extended North and South. The occupation of Czechoslovakia rendered that idea utopian for the time being.

Attempts to develop economic relations with as many countries as possible and the foreign policy of an uncommitted nation enabled Yugoslavia to establish trade with 120 countries. Trade is not only geographically dispersed, it is also diversified in terms of products exchanged. As a result a theory of "capillary trade" emerged. V. Pertot argues that small quantities reduce marketing difficulties, and S. Obradović adds that highly diversified trade reduces risks of business fluctuations. Empirical research lends some support to this hypothesis. P. Mihajlović finds that the concentrated pre-war export was very much dependent on external business fluctuations, while no such dependence appears to exist after the war (Mihajlović and Tanović, 1959, p. 77). Capillary trade also has its drawbacks. Obradović points out that it increases marketing costs and quotes approvingly Bičanić, who maintains that export concentration is a precondition for a permanent export position on the world market (Obradović, 1962).

Fast growth after 1955 led to profound structural changes. The share of exports of commodities and services in social product increased from about 13 to about 20 percent. The Yugoslav share in world trade doubled, but being still less than one percent, provides a justification for the capillarity theory. The share of those three traditional natural resource sectors in exports has been reduced from 90 to 50 percent (Fabinc *et al.*, 1968a, p. 144). Raw materials and manufactured goods changed their places in the structure of exports (Guzina, 1950:6 in 1939 to 13:50 in 1968). The once self-sufficient peasant economy is now only a matter of historical interest. It has been replaced by a relatively open industrialized economy participating actively in development of the world market.

Prologue: Rigid central planning in the period 1945–1951 implied a state monopoly in foreign trade. The domestic market was completely cut off from the outside world. The rate of exchange was just an accounting device without economic meaning. Export and import trade were conducted at prescribed domestic prices. The Fund for Price Equalization, created in 1946, compensated exporters for the differences between the domestic and export prices. Each transaction implied a separate foreign exchange rate. That was consistent with the principle of profitability at all cost applied in the home market. Exporters were obliged to surrender their foreign exchange proceeds to the National Bank which, in turn, supplied importers with what they needed. Foreign trade enterprises acted as agents for the Ministry of Foreign Trade and were obliged to implement import and export plans. Plans were defined in physical terms and so traders were not interested in prices and other trading conditions. The system was simple and consistent, but not very efficient. Yet, in the turbulent post-war years it did the job it was designed for.

The most important event in those years was the Cominform economic boycott. At that time details about operations of mixed Soviet–Yugoslav companies became publicly known and stirred great indignation. A certain number of these companies were created with a proclaimed aim of helping to develop the country. Capital was invested in even shares, profit was divided evenly, the Russians appointed their own people as general managers, insisted on preferential treatment and objected to Yugoslav financial control. All this reminded people too much of their pre-war experience with foreign capital and mixed companies were gradually liquidated. But the problem was more complex than that; economic relations among socialist countries were at stake.

In an interesting 1949 article M. Popović, then a member of the government and now the President of the Federal Assembly, explained the position that had been taken (1949). If a less developed and more developed country meet in the world market, they will exchange commodities with different labor contents. The more productive country will get back more labor than it gives away. This implies exploitation. Further, if in mixed companies profit is divided according to capital invested, a principle of distribution alien to socialism is introduced and as a result exploitation appears in yet another form. "According to socialist principles"—said Popović—"the entire surplus value, i.e., the entire profit obtained by the society after it had sold the commodity in the world market, belongs to the proletariat which has created that value . . ." (1949, p. 108).

To such theories, and not quite unexpectedly, Russian negotiators reacted rather laconically: "Torgovlja—torgovlja, a družba—družba" (trade is trade and friendship is friendship). But for Yugoslavia, then a year or two after the Revolution, socialism meant immensely more than trade; to put the two on an equal footing was profoundly shocking. Economic relations among socialist countries were seen as similar to the relations of the various regions within one country. Developed socialist countries had an obligation to grant aid to the less developed ones in order to speed up their growth and enable them to reach the same level of development in the shortest possible time (Obradović, 1962, p. 39; M. Popović, 1949, p. 70).

These were not abstract ideas; they were applied in relation to Albania. Yugoslav and Albanian partisans fought together during the war and relations between the two countries were very close. As a more developed country, Yugoslavia sent experts and material supplies to Albania. Tariffs were abolished and monetary units were given the same nominal value. Attempts to design a single system of prices failed because productivity differences between the two countries were too great. But they then continued to trade at their internal prices which meant that Albania exported at Albanian prices and imported at Yugoslav prices (the latter were somewhat lower than the Albanian on the average). In this substitution of world market prices by respective domestic prices. Popović saw the elimination of the exploitation characteristic of the world market mechanism (M. Popović, 1949, p. 128). In fact, however, this conclusion does not necessarily follow. To find out whether and how much Albania gained, one would have to calculate the entire trade in Albanian, Yugoslav and world prices and compare the value aggregates. And in order to make exchange equivalent in labor terms one would have to apply input–output analysis. Another policy measure had much more obvious implications. Albania was granted interest-free loans for an unspecified length of time. This was an early anticipation of the now familiar aid programs for underdeveloped countries.

Bulgaria was another country with which

Yugoslavia expected to eliminate tariffs and possibly even form a confederation. Yugoslavia waived Bulgarian reparations obligations for war damages, and after the Bled agreement in 1947, hopes went high in both countries. A few months later Stalin launched his attack, and soon all achievements were forfeited, all hopes buried. Former friends became enemies.

The Cominform economic boycott and the need to finance the Five-Year Plan compelled Yugoslavia to establish contacts with the world capital market. Ideological reasons and unpleasant experience with Western capital before the war and with Soviet capital afterwards made joint stock companies and mixed companies an undesirable form of import of foreign capital. Loans remained the only available alternative. But loans may also affect the economic and political independence of the country unfavorably. In order to prevent this from happening, V. Guzina suggested, in a paper representing the common opinion of the time, that foreign trade be conducted according to the economic plan, and a specified volume and structure of exports be secured (1950, p. 71). Guzina also held that autarchy was both impossible and undesirable, and favored development of an open but controlled socialist economy. These ideas were characteristic of foreign trade policy in the next decade.

Three Steps Towards Free Trade

By the middle of 1951 the new economic thinking reached the sector of foreign trade. As usual, market experimentation began with agricultural products. Exporters of certain agricultural commodities were allowed to sell their foreign exchange proceeds at a price which was obtained by multiplying the official rate by the factor 7. This foreshadowed the new official rate determined on January 1, 1952 at 1$=300 din (the old rate was 1$=50 din). Exporters were granted a retention quota of 50 percent with which they could finance imports of their own choice and sell imported commodities at free prices.

The transition from complete state monopoly to a system of free trade was not a simple affair. Various alternatives were discussed. In an important article early in 1952, D. Avramović, now a staff member of the World Bank, argued that a fixed exchange rate and, in particular, its exclusive use, cannot be practiced in a socialist economy. In order to secure the minimum volume and the necessary structure of exports and imports consistent with production and investment targets, the fixed exchange rate should be replaced by either physical allocation of goods or a system of multiple exchange rates. The latter is more consistent with a socialist market economy. Since foreign prices constantly fluctuate and since a full employment high rate of growth economy needs stability, there ought to be an Equalization Fund to absorb violent fluctuations. Thus, not only is there a need for multiple exchange rates, but these rates should also fluctuate. The capitalist principle of a fixed exchange rate *cum* business fluctuations must be replaced by a socialist principle of multiple fluctuating exchange rates *cum* economic stability and growth (Avramović, 1952).

Most of these ideas were soon tried out. In July of the same year the system of 17 price equalization coefficients was set in operation. Coefficients, applied to export prices calculated at the official exchange rate, ranged from 0.8 (for exports of agricultural products) to 4.0. Low coefficients were applied to imports of equipment and raw materials in order to keep their prices low.

A high degree of liberalization was envisaged in foreign trade, but in comparison to the liberalization of the home market, the liberalization of the foreign trade system proved to be a much tougher job. First of all—and again in contrast to the home market—the price of foreign exchange was set too low. Already in 1951 the actual average export exchange rate was 354 dinars for one dollar, and in 1952 it increased to 585 dinars which was almost twice as much as the official rate. The average import exchange rate was lagging behind appreciably (1$=440 din). Cheap imports exerted pressure on the balance of payments. The foreign exchange reserve of 4 percent of imports made economic interventions impossible. No wonder that the newly created foreign exchange market, DOM (Foreign Exchange Accounting Place), did not work. At first, exporters were obliged to sell only 55 percent of their foreign exchange to the Bank; the remaining 45 percent, representing their retention quota, could be used for imports of their own choice or sold to importers at the DOM. Already in October the retention quota was lowered to 20 percent, and that meant the death sentence for DOM. In the next year DOM rates soared to a level 6.8 times higher than the official rate. Average actual exchange rates went up as well.

In 1954 a series of desperate attempts was made to save the system. The accounting exchange rate was increased to 632 dinars for a dollar. Coefficients were revised and applied to DOM rates, and not to the official rate. A steep tax on the gains at DOM was introduced. A number of other complicated procedures were applied. DOM rates were brought close to the new accounting rate, which the authorities aimed for. Yet importers of raw materials could not compete any more at DOM for foreign exchange and so separate sales were organized for them. This reduced the amount of available free foreign exchange to something like one percent of the demand. The retention quota was reduced to only one percent. Prices of foreign exchange soared and by 1960 reached a level 12.3 times as high as the official rate. The National Bank replaced exporters as the only seller of foreign exchange (Mrkušić, 1963, pp. 301–315).

The first free trade attempt failed because the initial price for foreign exchange was set too low, initial reserves were too small, the share of the free market in foreign exchange supply was too small and disparities between home and foreign prices too great. It would have been rather difficult to find elsewhere in the world such relative prices, remarked V. Meichsner, as existed in Yugoslavia in 1955: one typewriter ribbon (2.800 din) equals a pair of shoes equals two yards of woolen fabric equals one third of an average employee's salary equals two-day full board in a first class hotel in a tourist resort equals 56 haircuts equals the monthly rent of a five-room apartment (Majhsner, 1956, p. 193). At that time three different foreign exchange regimes coexisted: the official rate, the regular and the separate DOM rates. Meichsner suggested that the number of coefficients be gradually reduced to only two, one for industrial and one for agricultural products. In 1957 M. Frković calculated deviations of actual exchange rates of various product groups from the average actual rate of 1$=779 din. It turned out that industrial exports and food, equipment and invisible imports were subsidized at rates between 21 and 35 percent, that there were export taxes between 16 and 21 percent for agricultural, wood and invisible exports and a protection rate of 105 percent for consumer goods imports (Frković, 1957).

By 1960 it had become clear that the foreign trade system needed a thorough revision. D. Čehovin evaluated the situation in three points. Enterprises were stimulated to press for an increase in coefficients, not to compete in the world market. Coefficients had ceased to be passive equalization instruments and were in fact transformed into active devices for increasing price disparities. Profitability calculations were made practically impossible

(Čehovin, 1960, p. 125). Mrkušić noted that in an economy where exports are price elastic and imports are not, there will be a constant tendency for export exchange rates to move away from the import ones. That required physical restrictions on imports (Mrkušić, 1963, p. 297). Both did happen. Higher export exchange rates were bound to produce inflationary pressure—via the money supply—as Avramović had already warned (1952, p. 24).

The recession that started in 1960 made things worse and stimulated the authorities to undertake a reform in 1961. This time an ample supply of foreign exchange was secured by foreign loans. But the other two mistakes of the 1952 reform were committed again: the new accounting rate was set too low (750 dinars for one dollar); the actual export rate in 1960 was 981 din and in 1961 went up to 1021 dinars (O. Kovač, 1966) and price disparities were corrected in only a few cases.

The strategy of the reform can be described as follows. Multiple exchange rates were abandoned and coefficients were replaced by a customs tariff. Instead of exchange rates varying between din 500 and din 1200 for a dollar, there was to be a single 750 rate with no protection for agriculture and lumbering, with 10–40 percent protection for consumer goods and 17–60 percent protection for equipment and other industrial products. Export was free and was supported by premiums and tax reductions. Exporters were supposed to sell foreign exchange to the National Bank but in most cases could buy back 7 percent of the amount sold for their own needs. About one fifth of imports was liberalized, and for the rest commodity quotas or foreign exchange allocations applied.

The deficiencies of such a strategy soon became apparent. Exports were retarded, imports accelerated. In order to keep the balance of payments deficit under control, import restrictions were multiplied and in 1964 the tariff protection was increased from 20 to 23 percent. Exports were stimulated by making foreign exchange allocation conditional upon export sales. Export premiums and tax reductions were rapidly expanding. Soon the old system of multiple exchange rates reappeared with all its inefficiencies (Institut, 1964).

The situation was worsened by the fact that about one half of Yugoslav foreign trade was oriented towards clearing currency countries, most of it towards COMECON. Both import and export flows with the COMECON countries are much more unstable than with the convertible market (Madžar, 1968). Both import and export prices on the COMECON market are higher than on the world market. Besides, it is easy to export to this market but difficult to import from it and vice versa for the convertible currency market. As there was one single exchange rate for both markets, the consequences should be obvious. Importers were oriented towards convertible currency countries, exporters towards clearing currency markets. The balance of payments deficit with the former increased rapidly, while there was an unabsorbed surplus on the trading account with the latter. A boom in 1964 produced unbearable pressure on the balance of payments. In the same year the cycle was reversed. The recession helped to induce the authorities to undertake another reform in 1965.

This time the structure of domestic prices was radically readjusted as explained in section 10. The actual export rate of exchange in 1964 was 1050 din; it was expected to increase in 1965 to 1200 din and the new official rate was determined at 1$=1250 din. Thus two fatal mistakes of two preceding reforms were avoided.

An additional element in the strategy consisted in the lowering of tariff protection from 23.3 percent to 10.5 percent with the traditional differentiation of rates from 5 percent for primary commodities to 21 percent for consumer goods (Domandžič, 1966). The necessary supply of foreign exchange was secured through the cooperation of the International Monetary Fund (IMF).

The ambitions of the reform were great D. Anakioski, one of the directors in the Federal Planning Bureau, describes the objectives of the reform as follows. The Yugoslav economy was to be integrated into the world market. Trade was to be gradually liberalized and the dinar made convertible. Exports were to rise relative to imports which would permit building up substantial foreign exchange reserves. The balance of payments deficit was to be eliminated (Anakiovski, 1969).

The new foreign trade regime became operative in 1967. About one quarter of imports was liberalized and retention quotas remained in most cases at 7 percent. For the rest there was a complicated system of inducements and restrictions; in order to achieve a proper regional distribution of trade, a category of imports from the convertible area was made conditional upon the purchase of a specified amount of clearing currency (Saveana, 1966). Export premiums and tax subsidies were abolished. Tight money policy was to keep prices stable, reduce internal demand and compel enterprises to export.

Once again the new regime failed to produce the results expected. After an initial burst of exports and a contraction of imports which in 1965 produced a small balance of payments surplus, imports began to expand faster than exports. Internal demand was checked, but so were exports. A balance of payments deficit reappeared and was increasing. Unpleasant clearing currency surpluses were cumulated. Import restrictions were multiplied. Export inducements were reintroduced. Differential exchange rates were back. The dinar was stable on the tourist market—dinar notes could be bought at rates close to the official one at all foreign exchanges—but a quiet devaluation was proceeding under the surface. None of the objectives quoted by Anakioski was achieved.

The ways in which the free trade reforms were carried out did not indicate an impressive professional competence. But in this respect Yugoslavia is not unique in the present world. The most popular method of policy making seems to be the method of trial and error. It has its drawbacks but, if applied with sufficient persistance, it also produces useful results.

So far I have been examining deficiencies. Let me now briefly evaluate the results. Since 1952 the span between extreme actual (resulting from actual revenues of exporters and actual payments of importers) exchange rates has been considerably narrowed. Actual exchange rates have become considerably more stable. The positive difference between the actual export and the actual import exchange rates of 303 din. in 1955 was transformed into a negative difference of 100 din in 1967. Government interventions in foreign trade operations have been reduced in every respect. About one fifth to one fourth of imports is firmly and completely liberalized either directly or via retention quotas and other arrangements. The tourist dinar is a stable and convertible currency. The stage is set for the last—if there is such a thing in economics—assault on free trade and convertibility.

The What-to-do-Next Controversy

The misfortunes of the third reform were not entirely unexpected. Mrkušić, A. Čičin-Šain and other economists evaluated various government objectives as unattainable given the policy pursued.

Soon a lively discussion developed focussing on three themes: protection, the nature of exchange rates, and convertibility.

I. Fabinc argued that every protection policy ought to be associated with a development program. Developing countries encounter serious bottlenecks in output capacities and shortages in material and financial means. Therefore, unlike developed countries whose protection policy aims at changing the structure of prices and incomes, developing countries must have a protection policy oriented towards changes in the structure of production. The main objective of tariff policy is to protect national production by producing a desirable differentiation of internal prices as compared with prices on the world market. There are, however, three important tasks which a tariff policy cannot perform. It cannot regulate the volume of imports, it cannot achieve the desirable structure of imports and it cannot regulate a regional distribution of trade (Fabinc, 1963, 1968b). One has to find other devices to do these three jobs.

Evidently, administrative interventions of the government are one possible alternative. It is, however, not acceptable as a dominant alternative in the Yugoslav setting. Next, a proper exchange rate system could do at least part of the job. This system could be based on one single rate, or on multiple rates, and the rate or rates could be pegged or be fluctuating. Out of these elements four main combinations and a number of variations may be formed. On the one extreme there will be a single pegged rate and on the other fluctuating multiple rates.

In the debate the Institute of Foreign Trade noticed an inconsistency in the traditional approach. The policy of a single rate usually imposes the elimination of multiple rates on the export side, while on the import side they are retained in the form of a customs tariff. In fact, however, the economic justification for multiple rates is the same for both components of foreign trade (Institut, 1964, p. 75). Mrkušić and O. Kovač of the IES suggested that the pegged rate be made flexible by the application of exchange rate ingredients such as tax reductions, preferential transportation rates and the like. But they find direct export subsidies unacceptable, presumably because they fear a proliferation of arbitrary government interventions (Bilandžić, 1967, p. 34). As far as the import side is concerned, Fabinc noted that fixed customs rates do not prevent their flexible application (by an appropriate definition of the customs value or by introducing point clauses) (Fabinc, 1963, p. 38). Other devices—such as a customs registration tax—are available as well. Thus even if the single fixed exchange rate is chosen as a basis for the system, the prevailing expert opinion favors making it flexible in both senses: it ought to be changeable in time and differentiable with respect to the fixed standard. The justification for this approach had already been provided by Avramović in the cited paper of 1952: a planned economy cannot tolerate that outside economic conditions and fluctuations be automatically transmitted to the internal market. This was now reiterated by U. Dujšin, who advocated not only flexible, but also fluctuating rates (Dujšin, 1968, p. 593). Mrkušić pointed out that if one wanted to keep the balance of payments in equilibrium either the exchange rate or internal prices will have to be continually adjusted. Since internal stability is obviously the first priority, the flexibility of the external value of money follows as a natural consequence (Mrkušić, 1967).

The government chose to base its policy on the pegged rate. This decision now came under attack. A pegged rate implied government interventions, which were resented. Fluctuating rates involved risks of instability, which the government was not willing to assume. Čičin–Šain thought that these risks could not be so great, that fluctuating rates required much smaller reserves and much less stringent conditions in terms of financial discipline, organization of the market etc. (Čičin–Šain, 1968b). A few years earlier, G. Macesich, an American economist of Yugoslav extraction, also argued in favor of fluctuating rates. He believed that "such a system would serve to integrate the country's economy more effectively with the world's economy by quickly indicating to planners when mistakes in the planning have been made. The correction of mistakes would not have to depend on intermittent changes in rigid official exchange rates" (Macesich, 1964, p. 202).

On the other hand Mrkušić argued that fluctuating exchange rates would generate speculation and would be destabilizing. He cited the Canadian twelve–year experimentation with fluctuating rates which he claimed ended with trade restrictions for about one half of imports (Mrkušić, 1969). Čičin–Šain suggested that speculation could be avoided if enterprises were obliged to sell foreign exchange as soon as it was earned. Capital movements would clearly require separate control.

Fluctuating exchange rates implied the existence of a foreign exchange market. The government feared that this might mean repeating the failures of the DOM. On the other hand enterprises and business chambers were pressing for higher retention quotas. The prevailing export opinion seemed to be in favor of the market, even if not for all currencies. Since the country ran a chronic surplus on its trading account with the clearing area as a whole and with most individual clearing currency countries, it seemed advisable to start market operations with these currencies (Bilandžič, 1967, p. 34). That would mean fluctuating rates for about one half of the foreign exchange proceeds. The next phase might be trading in convertible currencies, and finally a proclamation of the external convertibility of the dinar.

Čičin-Šain examined the pros and cons of approaching full convertibility via external convertibility, i.e., by satisfying Article VIII of the IMF agreement, or via internal liberalization. In favor of the former, he advanced the following three reasons: (1) the dinar might become a reserve currency, which would mean an interest–free credit for Yugoslav imports; (2) clearing countries might find it advisable to liquidate their clearing deficits in order to accumulate convertible dinar balances and (3) the financial prestige of Yugoslavia would increase. He felt, however, that these reasons were not particularly convincing. Even if fully convertible, the dinar would probably not be held as a reserve currency in any substantial amount, and in so far as clearing deficits were structural, they would not be remedied by financial devices. On the other hand, external convertibility would require substantial reserves and is the more difficult to achieve the higher the degree of internal liberalization (Čičin–Šain, 1967, 1968a). Liberalization would result in lower inventories—inventories are notoriously high in the Yugoslav economy—which would mean a considerable saving in foreign exchange and in working capital.

Later in the debate professional opinion swung in the direction of external convertibility. Mrkušić argued that in fact Yugoslavia maintained external convertibility with the convertible currency countries. If Yugoslav traders pay foreign exporters in their own currency, this is the same as if they paid in an externally convertible dinar. The official proclamation of external convertibility would lead to greater financial discipline, greater influence of the world market on internal costs of production and also to some foreign exchange economizing because foreign exporters would not insist on converting dinar balances immediately into their own currency (Mrkušić, in press). The Economic Institute in Zagreb pointed out that external convertibility would facilitate multilateralization of trade with the COMECON countries (Fabinc *et al.*, 1968a, p. 191).

As already noted, Yugoslavia belongs to neither of the trade areas in Europe and is politically uncommitted. As a result she encountered considerable difficulties in trade with her neighbors. However, why not transform this position of weakness into a position of strength? A country which went through underdevelopment, central planning and market organization and which is economically and politically uncommitted might perhaps become a desirable economic meeting place for three different worlds. If so, external convertibility is certainly one of the preconditions for making the mediating role of the Yugoslav market attractive for her partners from the West, the East and the Underdeveloped South (Čičin-Šain, 1968a, p. 82).

V. *Money, Banking and Public Finance*

Banking and Monetary Policy

There has been a lot of experimentation in the Yugoslav economy. This is true for the monetary field more than any other.

Banking can be organized in a centralized or decentralized fashion. Decentralization can be (1) regional, (2) functional or (3) both. Centralization can be (1) absolute or (2) partial. Thus there are five possible organizational solutions. All of them have been tried out at one time or another.

Banking for a Centrally Planned Economy: According to the Institute of Finance, in the socialist economy of 1949 money was a tool used by the state authorities to distribute social product in proportion to the labor of each working-man, to establish economic ties among enterprises and to exercise control over their activities. Money was also a means of accumulation and an instrument of control over plan fulfillment (Finansiski, 1949, p. 63). The banking system was expected to provide money which had such properties.

From pre-war times Yugoslavia inherited a certain number of private and state banks. The former were eliminated by 1947 and the latter were reorganized. The National Bank was a descendent of the Serbian National Bank created in 1883. The former State Mortgage Bank—the heir of a state bank set up in Serbia in 1862 (Uprava fondova)—continued to operate as the State Investment Bank. The Agricultural Bank of 1929 continued to operate in the same field. There was also a Handicraft Bank and, in view of ambitious industralization programs, it appeared advisable to set up a separate Industrial Bank.

The war had not yet been ended when a process of creating regional banks began: six republics—six regional banks.

For a country aiming at central planning, all these banks did not represent a very purposeful arrangement. In September 1946 a consolidation of the banking system began. All existing banks were merged into the National Bank, entrusted with short-term transactions, and the State Investment Bank, which was to deal with investments and foreign loans. Apart from dealing with short-term credit, the National Bank issued currency, performed general banking and agency services for the government and served as a clearing house for the entire economy. In 1948 the two-bank system seemed overly centralized. Since local enterprises and agricultural co-operatives played special roles at that time, 89 Communal Banks and 6 regional State Banks for Lending to Agricultural Co-operatives were formed. Communal banks were universal banks: they were for servicing local budgets, extending short and long-term credits, collecting savings, controlling plan fulfillments of local enterprises. Banks charged a one percent interest rate which was in fact a commission charge for their services. It was not deemed appropriate for a socialist system to charge interest as a price for capital.

Since it is much easier to control financial transactions conducted via bank accounts than those made in cash, already in 1945 all enterprises and other non-private transactors were obliged to have drawing accounts with the bank. Soon about nine-tenths of payments were conducted without using cash. This was one of the lasting results of the early period of banking development. Payments through bank balances developed into a unique internal payment system, channeled through local offices of the National Bank. It embraced all banks, post offices, enterprises, government funds and a considerable part of the private sector and connected all money streams of the economy into a single consistent system (Vučković, 1963, p. 366).

In many respects the early Yugoslav monetary system was a replica of the Soviet model. This is particularly true for the three instruments of monetary control: credit planning, cash distribution and the automatic collection of invoices.

Credit planning was the only instrument to survive the administrative phase. Until 1950 credit planning simply meant summing up the credit needs of individual enterprises. This was done by planning authorities. The bank was supposed to implement such plans in a routine way. Later, credit plans were transformed into credit balances, which meant that needs were balanced with means. Banks were made responsible for drawing up credit balances (Vučković, 1956, p. 172). The planned amount of credit for individual enterprises was obtained by dividing the output target into an individual capital-turnover coefficient and then subtracting the enterprise's working capital (Vučković, 1963, p. 366).

The main purpose of cash distribution plans was to control recipts (mainly in retail trade, catering and passenger transport) and expenditures (primarily for wages and payments to peasants) made in cash (Stevanović, 1954, pp. 145–46). The cash plan was made for territorial units and for separate money streams and so provided useful information about receipts and expenditures of the population and about various channels in which the money was circulating in the economy. But it was a rather rigid instrument with not much use outside central planning and was therefore abandoned in 1951.

In order to enhance financial discipline, enterprises were forbidden to grant trade credits to each other. The automatic collection of invoices served the same purpose. The bank would automatically credit the seller's account when goods were shipped and then charge the buyer's account. In this way no mutual crediting could be practised. Payments were carried out smoothly. If there was no money in the buyer's account, credit was automatically extended. This, of course, meant that credits would expand beyond the limits set by the credit plan. At first, such matters did not worry planners too much; physical targets and not money flows were important. Other consequences were more disquieting. The total volume of credits depended more on debtors then on banks. The necessary discipline was jeopardized. Sellers did not care about the solvency of their buyers, and also tended not to pay sufficient attention to delivery terms, assortment and quality of goods. Buyers did not mind accumulating excessive inventories. After a while careless buyers had to be put on "black lists," their drawing accounts were blocked and in many cases they were brought before the courts. The automatic payments mechanism broke down and was in 1951 replaced by free contracts among the trading partners (Vučković, 1957, p. 21).

Learning by Doing: What sort of bank ing system was appropriate for a self–management economy? Centralized or Decentralized? There was a lively discussion on that issue. E. Neuberger surveyed the principal arguments advanced in favor of the one or the other alternative (1959a). Whatever the merits of these arguments might have been per se, the government decided to play safe. No one could be sure of the business behavior of labor–managed enterprises. It seemed advisable that decentralization in the market for goods and services be accompanied by strict centralization in the financial sphere. All other control instruments, remarked J. Pokorn, were to be replaced by bank control and supervision (1956). In March 1952 Communal banks ceased to exist and other banks were merged with the National Bank into one single giant bank with 550 offices and 16,000 employees.

In order to make control as efficient as possible, the working capital of enterprises was transferred to the Bank. Enterprises were to pay a reasonable interest rate, which was to induce them to economize with the credit money.

The shorter-term interest rate was differentiated according to turnover velocity of working capital and ranged from two percent for crude oil production and agriculture to 7 percent for electric power plants. This span was reduced to 5–7 percent in 1953. It was again increased and the rates differentiated in a somewhat different way, for different kinds of credits, in 1954. Experimentation with interest rates continued even later, and in 1956 there were 25 categories of active interest rates (Vučković, 1957, p. 183).

The so–called "social accounting" represented one lasting result of the 1952 reform. The Bank established special accounts—at first thirteen of them—for all important transactions of each enterprise. All changes that took place in the current account of an enterprise were entered here. In this way: the Bank and the government had up–to–date information; the Bank was able to exert stringent control—it would stop any irregular payment, which was particularly important for payments related to wages; the Bank checked the fulfillment of tax and other obligations of an enterprise towards the state. The system was later simplified, the number of separate accounts was gradually reduced and the Bank began to rely more on quarterly accounting statements by the enterprises. A standard accounting scheme, obligatory for all enterprises, made this task a routine matter. In 1959 the social accounting with its drawing accounts system for the entire economy was separated from the National Bank and turned into an independent social service. The work was computerized and the service became very efficient. A little later it was discovered that the Social Accounting Service's monopoly on the payments traffic was not an obstacle to enterprises keeping their financial resources with the banks of their own choice. Today every non–private income earner has a drawing account with the Social Accounting Service, and pays commission charges, and at the same time has a deposit account with one of the banks, and receives interest on deposits.

The proper procedure to be used in extending short–term credits was one of the important problems the all–embracing National Bank had to solve. In those days of romantic beliefs in the possibilities of inventing simple problem-solving devices—such as the Rate of Accumulation and Funds—that would eliminate the arbitrariness of a bureaucratic apparatus, the Bank hired a couple of mathematicians and asked them to invent appropriate formulae for credit extension. A booklet with several dozens of such formulae was published in 1952 (Miljanić *et al.*, 1956). They were based on turnover coefficients of credits and ratios of sales to costs. Since parameters to be used in formulae could be calculated only as some sort of averages, it was soon discovered that some enterprises got some more credits than they needed, while others badly lacked the money to keep production going. Formulae were abandoned and in 1953 the amount of credit extended was related to the maximum quarterly credit used by the enterprise in the previous year. This favored last-year debtors and penalized good entrepreneurs and had to be abandoned. But the idea of some automatic credit evaluating mechanism was not abandoned.

In 1954 the Bank experimented with credit auctions. Vučković explained that credit auctions were to be a kind of socialist credit market where the supply and the demand of money would meet and determine the general conditions for credit extension (1957, p. 38). The Bank expected that less profitable enterprises would refrain from asking for credit because they would not be able to bear high interest rates. It turned out that precisely the less profitable or unprofitable enterprises were prepared to offer the highest interest rates—up to 17 percent—because they considered credit the only available solution for their problems. The Bank then set the marginal interest rate at 7.5 percent. But this was a negation of the whole idea of auctions. Soon credit was extended automatically to every enterprise that had satisfied the formal conditions of an auction. Since all automatic devices provde inefficient, in 1955 the Bank fell back on the traditional banking practice of an individual evaluation of every credit request.

By 1954 two facts were established: (1) the NES worked well as a whole but (2) the centralized bank left much to be desired. As soon as that had become clear, regional and functional decentralization were initiated. One of the main justifications for decentralization was the socio-economic incongruity between self-management in the commodity market and state monopoly in the financial market. Vučković quoted approvingly the governor of the National Bank, who declared that in a decentralized banking system the credit function would be subject to the control of social self-government instead of bureaucratic management (Vučković, 1957, p. 86). Communal banks with all their diverse activities were re-established. The banks were obliged to keep reserves with the National Bank of up to 30 percent of demand deposits and 100 percent of investment funds. In the next three years three specialized federal banks were added: a foreign trade bank, an investment bank and an agricultural bank. The National Bank was relieved of investment and some other banking operations. Each bank was run by a managing board whose members were partly appointed by the authority that founded the bank and partly elected by the bank's personnel in the proportion 2:1.

After all these changes had taken place, it appeared appropriate to give back working capital to enterprises. This was done in 1956, and the system was stabilized for the time being. Working capital was not given back free of charge; enterprises were obliged to pay an interest rate of six percent.

Banking for a Self-Government Economy: It took eight years before a formerly administratively run economy learned how to handle a few basic financial mechanisms. The task of creating an adequate institutional system in the financial sphere was yet to be accomplished. It took eight more years before an outline of such a system became visible.

The deficiencies of the banking system as it developed until 1960 were described by V. Holjevac as follows (1967a). The National Bank offices were inefficient, unimaginative, engaged in distributing the planned increase in credits and executing the decisions of the head office. Communal banks fell under the complete control of local authorities which often made it impossible to conduct a sound business policy of profitable and safe investments. The federal government often directly interfered with the banking business by immobilizing certain kinds of deposits or by running a deficit inconsistent with the social plan. In order to overcome these deficiencies a series of reforms was undertaken. As in the post–1952 period, reforms were carried out in two–year intervals starting with 1961.

In 1961 communal banks became basic and universal credit institutions. In order to eliminate the monopolistic influence of political authorities, a two–thirds majority of members of the banks' managing boards were nominated by workers' councils of the enterprises located in the territory of the bank. Next, eight regional banks reappeared. They were to serve as mediators between communal banks—which were required to keep a 5 percent reserve with respective regional banks—and the National Bank. That was a rather unfortunate arrangement, since it caused the disintegration of the national credit market into six regional markets with different business conditions etc. (Miljanić, 1964, p. 53). This mistake was rectified four years later.

In 1952 an interesting new institution was created. It was called Joint Reserve Funds of the Enterprises. D. Dimitrijević describes Joint Reserves as a semi–financial intermediary. Joint Reserves—created at communal and republican levels—grant credits to those enterprises which have losses, are not competitive, have an unsound financial position and are not eligible for regular bank credit (Dimitrijević, 1968a, p. 19).

For more than a decade Yugoslav banking practice, and monetary theory, maintained a fundamental difference between fixed capital and working capital financing. This made sense in a centrally planned economy, but led to mistaken policies in a market setting. It was now realised that working capital was not homogeneous: it consisted of a constant part, which could and should be financed as fixed capital, and a fluctuating part which was a proper object of short–term credits. In 1961 enterprises consolidated the fixed capital and working capital funds into one single business fund. Thus all liquid assets could be used both for current payments and for capital formation.[8] In order to increase the financial independence of enterprises, they were encouraged to finance the constant part of the working capital out of their own funds and to rely on bank credit for the fluctuating part only. But that was not enough for a full–fledged credit policy.

Policy makers had to solve the following problem: design a flexible credit policy with a minimum of administrative allocations when there is no proper money and capital market. They decided to use so–called qualitative control, which implied regulating the demand for credit. The new policy was introduced in 1963 and one of its architects, N. Miljanić, governor of the National Bank, gave a detailed account of it in a book published a year later (1964). According to Miljanić, final demand ought to be financed out of income produced. This implies that inventory formation should be financed out of accumulation. The Governmental budget deficit could be used as a source of new money, but that is not desirable because in the absence of a money market, the distribution of such money occurs in a haphazard way and cannot be controlled. Miljanić even insisted that the federal budget should be balanced in any case (1964, p. 31). This contention, though clearly not defensible in theory, has some justification in practice in view of the sometimes less then responsible deficit financing of government agencies. The official document of the National Bank adds that in case of a recession it is preferable to increase, selectively, the money supply rather than to run a budgetary deficit (Narodna banka, 1965, p. 28). Since the liquidity trap is non-existent in the Yugoslav economy, this is a valid statement.

---

[8] However, enterprises were still obliged to hold five separate accounts, apart from the drawing account, with the Social Accounting Service. These accounts (depreciation, undistributed profits, non–business expenditures, and two types of reserves) were operated under special rules designed to induce enterprises to behave in a proper business fashion (Miljanić, 1966). Separate accounts, of course, reduced the possibility of rational use of money, since it could not be freely transferred from one account to another. However, gradually separate accounts have been eliminated.

New money ought to be used to finance primarily the circulation of commodities. Thus credit is given on the basis of some evidence, invoice or bill of exchange, that a commodity has been sold by a producer or bought by a merchant. Credit cannot be given for sales to final buyers (government, investors, consumers). As exceptions to the rule and on the basis of individual evaluation by the bank, credits can also be given for seasonal stocks and for stocks due to some circumstances beyond the enterprise's control. (In fact, credit for stocks, far from being an exception almost reached the level of credits for commodity circulation) (Miljanić, 1964, p. 72). Apart from this first category of credits, which creates some sort of neutral money, credits can also play an active role in supporting production. Such are credits for specific ventures, primarily for exports, agricultural production and for building apartments for sale. Miljanić also noticed one difficulty with his system. Business operations require that an enterprise always have at its disposal a certain amount of money pure and simple. This money is a part of constant working capital, but, being money, should not be financed out of income. On the other hand, if it is financed by credits, they are clearly not short-term ones. Miljanić feels that revolving credits might do the job (1964, p. 88).

This system lasted for four years and produced some good results. Enterprises knew in advance what conditions they must fulfill in order to obtain credit from the bank. Commercial banks were sure to get credits from the National Bank if they fulfilled the prescribed conditions. But the system was also deficient in many ways. B. Mijović, a director in the National Bank, pointed out that qualitative control (conditions, purpose, duration and kinds of credits) could not quite achieve the aims of quantitative regulation of the money supply. The National Bank had to generate a constant stream of detailed and extensive instructions, which became particularly cumbersome. Since not all practical cases could be envisaged and regulated in advance, the handling of borderline cases caused considerable difficulties. Frequent institutional changes elsewhere in the economy caused additional difficulties (Mijović, 1967, pp. 73, 112). By 1967 the credit system was ripe for a new reform. This time supply of—and not demand for—credit was made a primary object of monetary control. Selective control was accommodated within a system of quantitative regulation.

The three types of credits—investment, commercial and consumer—led to a law providing for the setting up of three types of banks: investment banks financing fixed and constant working capital, commercial banks extending short-term credit, and savings banks dealing with consumer credit. Table 9 summarizes the latest organizational changes (Basaraba, 1967, p. 78).

Organizational changes reflected very definite policy changes. (1) Federal, Republican and communal banks disappeared. All banks can in principle conduct their transactions over the entire territory of the country. This deterritorialization policy came as a response to frequent complaints against parochialism and unsound political pressures of local and republican authorities. (2) The market orientation of banks resulted in a concentration process that reduced the total number of banks by one half in only three years. By the end of 1968 the number of banks was further reduced to 74. This number ought to be compared with 700 private banks before the war. But the most important was (3), the change in the setting up and running of the banks. Here at last, a solution consistent with the organization of the rest of economy was found.

Banks are now established by enterprises and socio-political communities (federal, republican, local) as equal partners. In order to be independent business establishments, banks have their own capital, called the credit fund. The founders invest their capital in the credit fund of the bank and become shareholders. At least 25 founders are required for any bank so as to preserve the essentially service function of the bank. The bank is managed by enterprises and socio-political communities in proportion to the amount of their capital invested in the credit fund. Shareholders are entitled to dividends depending on business success. These dividends cannot be distributed in wages, but can only be used for capital formation. In order to prevent monopolization, no single shareholder can have more than ten percent of the total number of votes in a bank's assembly regardless of the amount of capital invested. Also no enterprise or socio-political community can be refused the right to invest in a bank and take part in its management. The Assembly of a bank consists of investors and representatives of the bank's personnel. It appoints the Executive Committee, the director and his deputy. The Executive Committee implements the bank's general business policy. The Credit Committee deals with individual requests for credit except in some special cases. In order to ensure an objective and expert business evaluation of requests, the Credit Committee is composed of the bank's own experts. The employees of a bank have their own self-management bodies which deal with the distribution of personal income, use of various funds, personnel matters and the like and, through representatives on the Executive Committee and in the Assembly, participate in the management of the bank.

TABLE 9.—BANKS IN YUGOSLAVIA

November 1964		June 1967	
Type of Bank	Number	Type of Bank	Number
Communal banks	206	Commercial banks	61
Republican investment banks	8	Mixed banks	39
Specialized federal banks	3	Investment banks	11
Total	217	Total	111

After a network of commercial banks had been established, the National Bank discontinued its direct business contacts with enterprises and became a central bank in the traditional sense.[9]

In its function of regulating the money

---

[9] Neuberger examined the role of central banking under three types of economic systems, the Yugoslav system before 1961 being one of them (1958).

supply, in 1961 the National Bank had the following weapons at its disposal (Golijanin, 1967, pp. 95–104).

1. Currency issue.
2. Sales of foreign exchange.
3. Fixing of terms for extension of short-term credit by communal banks.
4. Legally required reserves held by communal (later by commercial) banks with the National Bank. The upper limit was set at 35 percent of liquid deposits.
5. Limits for interest rates (in practice 8–12 percent).
6. Restriction of the use of certain kinds of deposits. This instrument was often and indiscriminately used, which greatly annoyed the owners of funds. I. Perišin points out that in the period 1954–1962 between 34 and 45 percent of total deposits were blocked in this way (1967).
7. Special credits extended by the National Bank to other banks. These credits were used to finance about one half of all short-term credits extended by commercial banks to their clients.
8. Consumer credit policy.
9. Consultations and recommendations.

Compared with traditional banking, some items appear superfluous, but one important item is missing: there is no place for an open market policy since there are, so far, no treasury bills. Instrument 7 is a substitute for that. By special credits new money is created and the liquidity of commercial banks ensured. If a bank wants to reduce excessive liquidity, in order to avoid paying passive interest, it can do so by repaying its credit to the National Bank.

As already mentioned, the 1967 reform replaced credit demand control by credit supply control, and so the functions of the National Bank had to be adjusted accordingly. Instruments 3 and 6 were abandoned and the existing amount of special credits was frozen and could not be increased. Several new instruments were added:

10. Rediscount credit, which is used as an instrument of both global and selective control. It amounts to about 12 percent of all commercial credits. In order to qualify for getting this type of credit a commercial bank must fulfill two conditions: (a) its total indebtedness with the National Bank cannot be greater than its demand deposits; (b) at least one-half of its short-term credits must consist of credits with repayment periods shorter than three months. Condition (b) is a special type of liquidity reserve requirement designed for the Yugoslav environment where there is enormous pressure to use short-term sources for investment loans.
11. Discount rate.
12. Quantitative restriction of credit as an exceptional measure.

This is an impressive array of weapons which, if inappropriately used, can cause considerable damage. In the section on monetary policy we will see how this can happen.

In 1967 a daily market was set up within the Association of Banks as a particular kind of stock exchange for supply and demand of short-term capital. Banks in need can obtain credit for a period not exceeding 15 days (Basaraba, 1967, p. 81). At the time these lines were written, the Federal Parliament passed a package of financial laws providing, among other things, that shares in a bank's capital can be sold to the business public, but not to socio-political communities and to banks themselves. These two events may be considered as proper beginnings of a stock exchange development—of the Yugoslav variety, of course.

Let me close this section with a note on monetary planning. On the basis of reliable and up-to-date information provided by the Social Accounting Service, a sophisticated system of flow-of-funds accounts was designed. Since 1967 this system has also been used for annual and monthly monetary planning, thus replacing the old-fashioned credit balances. Its author, Dimitrijević, gave a technical description of the methods used in his 1968b article.

Investment Financing: The amount of professional literature on investment financing varies in inverse proportion to the number of complaints against the state of affairs in this field. It is difficult to figure out why this is so. Perhaps it is because investment financing is in a sense a borderline case: neither monetary theorists nor fiscal policy experts nor predominently physical planners feel competent to deal with it. In any case investment financing has been one of the weakest links in economic policy for a long time, and yet no serious study of its problems has been undertaken so far. Thus I will confine the exposition to a description of actual development.

Capital formation may be financed by fiscal means, i.e. out of taxation, or out of enterprises' own funds, by bank loans or by means of securities of various kinds. This is roughly the order in which the various kinds of investment financing have been tried out in Yugoslavia.

Early in 1945 the government created the Fund for Reconstruction whose resources consisted of confiscated war profits[10] and of proceeds from sales of goods supplied by UNRRA. Very soon loans given by the Fund were written off and capital formation was financed in the budgetary fashion typical of a centrally planned economy. Investment resources were allocated by the plan and given to enterprises from the budget free of charge. Enterprises could not sell capital goods; they could only transfer them to other enterprises after having obtained permission to do so. Since the state was the only owner of capital and prices did not matter much anyway, this arrangement was consistent with the rest of the system.

The crucial year of 1952 inaugurated important changes. The Federal budget as a source of investment finance was replaced by the Fund for Basic Capital Development. Investment resources were still allocated without repayment obligations, but the creation of the Fund led to a division of the budget into two separate parts: one was related to administrative expenses and the other consisted of various investment and interventionist funds. This was to become a permanent feature of the Yugoslav budget.

In 1952 the federal government concentrated just about all investment resources in its Fund. That served the purpose of gaining time for the preparation of a more thoroughgoing reform. Already the next year Funds for Crediting Investment Activities were formed. Enterprises established their own investment funds financed out of profits that by the plan were left to them. Both measures led to a considerable decentralization of capital formation financing. The system assumed its more permanent shape in 1954 when Social Investment Funds (SIF) were created at all levels, federal, republic, district and communal. Since then, until the latest reforms, Social Investment Funds were granting loans to business enterprises, while capital formation in the non-business sector (schools, hospitals, government offices, etc.) continued to be financed out of the government budget. The creation of SIF—which tended to multiply as time went by—had an interesting behavioral consequence. Since all levels of the government were under constant heavy pressure to invest, and funds were separated from the budget, their resources tended to be inflated beyond anything envisaged by the Social Plan. In the period 1955–1960 the volume of investment surpassed the target established by the Social Plan by 20 percent (Vasić, 1963, p. 2157).

---

[10] In a similar setting after the First World War the government had great difficulties in introducing the tax on war profits and once the required law was promulgated, it could not be implemented (Milojević, 1925, pp. 168-82).

The reform of 1954 introduced two other important innovations. One consisted of the transfer of capital assets to enterprises. For the privilege of using social capital, they had to pay an interest rate of 6 percent, which was in 1965 lowered to 4 percent. Interest had to be paid on capital used regardless of the source of its finance. The proceeds from this interest as well as the repayments of the loans granted represented resources of the General Investment Fund operated by the federal government. The interest rate on social capital was differentiated according to the aims of price policy and according to the capital intensity of particular industry groups. It ranged from close to zero for agriculture to 1 percent for electric power generation and coal production, to 2 percent for transportation, to 4 percent for ferrous metallurgy and to 6 percent for most other industries. In this way the interest burden, as a percentage of net product, was more evenly distributed among various industry groups. The average rate of interest amounted in 1961 and 1966 to 2.8 and 1.3 percent respectively in terms of capital and to 3.8 and 2.4 percent respectively in terms of net products (Trklja, 1968, p. 23).

The second innovation is related to investment auctions. There are four types of investment allocation decisions: (1) the level of total investment, (2) the allocation of investment funds among sectors of the economy, (3) the allocation among firms within a sector, and (4) the allocation among technological variants within a firm (Neuberger, 1959b, p. 103). The last decision is made by the enterprise, while the first two are determined by the plan. After priorities have been determined, and investment allocated to the various industry groups, the allocation among the enterprises may be carried out by auctions.

This is an old textbook idea. In various texts on socialist economics with neoclassical background one can find statements that run roughly as follows: "In principle, the applicants would be listed according to the level of the rate of interest they offered and if two offered the same rate, the one who offered the shorter period for repayment of the loan would be given preference. The bank would go down the list until the amount allocated for this auction, or category within the auction, was exhausted, and the rate of interest offered by the first intramarginal applicant would become the one that everyone paid." In fact, this is not an invented quotation, but Neuberger's description of actual investment auctions in Yugoslavia (1959b, p. 93). In theory one could, of course, improve this scheme in various ways. One could apply price discrimination in order to siphon out all non–labor income contained in the difference between the offered and paid interest rate, or one could replace point offers by schedule offers. In practice the experiment did not achieve great success. It was soon discovered that the two price criteria—the interest rate and the repayment period—were insufficient. Thus other criteria were added: the percentage share of participation with own resources (differentiated according to industries and ranging from zero for electric power to 80 percent in manufacturing), the shortest period of construction, the lowest cost per unit of output, and regional effects (Vučković, 1963, p. 372; Hanžeković, 1967b, p.

TABLE 10.—THE COMPOSITION OF INVESTMENT IN FIXED CAPITAL BY SOURCE OF FINANCE, EXCLUDING PRIVATE INVESTMENT
(IN PERCENTAGES)

	1948	1951	1952	1953	1954	1955	1960	1962	1964	1966	1968
Social Funds and Budgets	99	98	98	87	74	64	52	59	36	16	16
Federation	60	50	95	71	50	47	37	30	7	6	9
Republics	27	41	2	11	12	9	7	9	8	3	3
Communes and Districts	12	7	1	5	12	8	18	20	21	7	4
Work Organizations	1	2	2	13	26	35	37	38	32	46	37
Business	1	2	2	13	26	27	31	30	26	39	31
Non-Business	—	—	—	—	—	8	6	8	6	7	6
Banks	—	—	—	—	—	1	1	3	32	39	47

Sources: For years 1948–1955: *Yugoslav Survey*, 1963, *15*, p. 2167.
For years 1960–1968: *Statistički bilten SDK*, 1969, *3*, pp. 68–69.

220). The main defects of auctions appeared to be the following ones. It takes time and it is very costly to prepare an application for credit. Auctions are held at widely spaced points of time which may not correspond with the enterprise's need for investment funds. As in the case of credit auctions, enterprises were ready to offer high rates of interest just to secure the loan. They did not worry too much about future repayments because the tradition of free social capital was still very much alive and because it looked obvious that a plant of any size cannot be closed down "just because the loan cannot be repaid." Thus the authorities in charge of SIF had to examine every case very thoroughly as they would have had to do even without auctions. According to Neuberger's estimates, at most one–third of all investments at any time were allocated through auctions. In such circumstances auctions gradually degenerated into an old–fashioned administrative distribution of investment from government funds.

Auctions failed. The criteria used for investment allocations from SIF had never been very transparent—another reason for the lack of analytical literature—and had always been greatly influenced by political considerations. As a result "political factories" appeared. All important investment projects were somehow multiplied six times, one for each republic. Besides, Social Investment Funds absorbed two–thirds of total investment resources, and owing to participation requirements, controlled directly an even larger share of total investment. Inefficiency and bureaucratic control were not quite compatible with the self–management aspirations of the economy. Enterprises pressed for an increase in their share in investment finance. The data on actual development in characteristic years are given in Table 10.

A considerable share of investment money in the SIF was obtained through taxation. When in 1962 these "contributions to Social Investment Funds" were raised by 50 percent (Vuksanović, 1966) there was a general outcry against the "expropriation." It was requested that "state capital" be done away with. Two years later the contributions to SIF were abolished, and the funds transferred to bank credit funds. That is why bank investment loans increased so sharply in 1964. The starting principle of the reform of 1965 was: to leave at the disposal of enterprises a larger share of their savings and consequently to restrict the role of socio–political communities in investment decisions (Jovanović, 1965, p. 3222). The pendulum was pushed a little too far in the decentralization direction because it was requested that even large capital intensive projects (power generation, communications) also be financed out of capital concentrated in banks.

The role of the Federation in investment was reduced to the operation of the Fund for Undeveloped Regions that would distribute annually to undeveloped regions close to two percent of national income as investment funds. Republican and communal funds also diminished considerably. But the share of enterprises, with the exception of a short-lived post-reform increase, remained stagnant. As Table 10 shows, what actually happened was that the Federation and the banks simply changed places in investment financing.

In a situation of chronic excess demand for investment resources, banks could easily assume a dominant role. The sum of the regular and penalty rates of interest could be as high as 18 percent. The first recession—which in fact followed the reform—was bound to reduce the investment funds of enterprises and make enterprises more dependent on banks. D. Vojnić points out that in 1968 the repayments of bank loans amounted to 111 percent of net profits of enterprises (1969, p. 89).

With almost one-half of investment resources under their control, banks established themselves as a dominant force in the investment market. What should be done to safeguard the independence of enterprises? The answer is by no means clear. The present discussion has concentrated on possible improvements of the capital market. In 1963 government bonds became negotiable and in 1968 the first enterprise bonds appeared. In 1969 bank shares were invented and the present author has suggested that participating debentures be introduced (1967b). The securities market could supply at least part of the capital outside the bankers' control. Pooling resources and joint ventures are encouraged. After the Social Investment Funds had been abolished, interest on social capital became a mere capital tax that flowed into the government budget. A political decision was taken to abolish this capital tax as soon as possible. It is now being suggested that this interest—amounting to about one-eighth of business investment—be given to enterprises as resources earmarked for investment (i.e., it would be treated similarly to depreciation funds). It will not be surprising if in a little while another reform in this field is carried out. After a money market has been to a certain extent adequately organized, its twin, the capital market, can surely not lag behind for very long.

Anti-Inflationary or (Anti-) Anticyclical Monetary Policy: In a centrally planned economy market disequilibria result in physical shortages; in a market economy they are reflected in inflation. The age-old discussion about the real causes of inflation was resurrected among Yugoslav economists, in particular after 1961.

Monetary theorists, not unexpectedly, tended to see the source of all troubles in an uncontrolled expansion of money supply. M. Ćirović argued that the increased commodity prices represented the way in which the economy adapted itself to an excessive expansion of credit and money supply (1966, p. 183). Similarly M. Vučković believed that inflation was essentially a product of excess demand. Since new money brings along new demand unaccompanied by supply, a market disequilibrium arises and generates increases in prices. The excessive expansion of short-term credits is a consequence of the following deplorable practices: short-term credit is used (irregularly of course) for long-term investments, for non-salable stocks, to cover losses, to finance budget deficits and to finance taxes at all levels of government (Vučković, 1967, pp. 128–29). The last mentioned practice is probably also one of the Yugoslav inventions in the field.

Owing to a fairly completely budgetary decentralization, local governments are very keen on squeezing out of "their" enterprises every possible dinar. In the early days of the NES they could do so by tailoring taxes so as to leave the coffers of the enterprises empty. This phenomenon had been described by Miljanić and Vučković already in 1956 (Miljanić, 1956; Vučković, 1956). Thus in 1954 in one single year, communes managed to increase their budgetary revenues by 98 percent (Vučković, 1956, p. 173). In order to comply with these patriotic requests, enterprises would have to increase prices or ask for credits or both. Credits were readily granted because paying taxes on time had always been considered a first priority. After the budget system had been somewhat more efficiently designed, the arbitrariness in taxation was reduced, but whenever in need communes would simply delay payments for goods and services they bought. In this respect republics and the federation have also been guilty until this very day. It is not surprising, therefore, that the business community does not trust their governments too much and tries to get rid of any "bureaucratic" control.

Now, though it is true that credit was excessive and money supply inconsistent with stable prices, it does not necessarily follow that prices were the consequence and credit the cause in the inflationary process. The hypothesis was tested in the IES and it turned out that there was either no correlation between credit and prices or there was a slight *negative* correlation: higher credits-lower prices. This paradox will become understandable in a moment.

Prices are predominantly determined by changes in wages, and so inflation is most of the time a cost push inflation. As already mentioned in the section on Price Policy, wages appear to be a function of capital intensity, technological rent and institutional monopoly (banks, insurance companies). Wage increases in privileged work-organizations initiate wage increases throughout the economy, and whenever prices cannot bear a cost increase, they are revised upwards. Bajt adds that the high degree of price control increases the pressure of excess demand on the free section of the market, which then generates price increases, and that inefficient investment planning produces an inadequate structure of output which in a semi-closed economy makes it difficult to match demand (Bajt, 1967a; Bogoev, 1967).

Business cycles complicate matters even further. Prices are formed in Yugoslavia in a rather peculiar way. Depreciation and interest on social capital represent fixed elements. Wages, as everywhere, are inflexible downwards. For reasons explained in the next chapter, all taxes are tied up with wages and vary proportionally to wages. Since tax payments enjoy high priority it may happen—and did happen—that the total amount of taxes collected increases in the trough of a depression. Finally, repayments of loans represent an additional fixed element. Thus, as soon as there is a slight retardation of production, the enterprise finds it impossible to cover costs and has to run losses,—or increase prices.

In a downswing a labor-managed enterprise will not dismiss workers. Thus production will be continued and inventories accumulated. Inventory accumulation is financed out of profits and credits. When these two sources are exhausted, involuntary trade credits and price increases will replace them. As far as inflationary pressure is concerned, we may expect price increases in the downswings and stable prices in the upswings. Figure 1 confirms such an expectation.

The analysis just sketched—a result of research of the IES—was unknown at the time monetary reforms were designed and implemented. The traditional view that inflation means "too much money chasing too few goods" gained wide acceptance. All one had to do, so it was thought, was to curtail the supply of money and the economy would be stabilized. Stabilization was envisaged exclusively as price stabilization. In the program of the 1965 reform employment targets were not even mentioned. Foreign exchange reform, membership in GATT and co-operation with the IMF were interpreted as an international obligation to keep the dinar stable at all costs—a task which even a Tory government would nonadays be reluctant to undertake, but which was cheerfully attempted in an economy innocently unaware of what it might mean. Tight money policy was to be the only device for achieving price stability. There were some doubts about the wisdom of such a policy, but critics were frowned upon and the policy was implemented. That proved to be fatal. Since prices vary inversely to the cycle, an anti-inflationary monetary policy meant an anti-anticyclical policy, a policy of continuous and direct destabilization.

The vicissitudes of monetary policy in the last eight years have been analyzed by Holjevac (1967b) and Perišin (1969), recently appointed Governor of the National Bank. I will mainly draw on their work and on the research conducted in the IES in the text that follows.

In 1960 the cycle reached the upper turning point (see Fig. 1). That passed unnoticed, but price increases were noticed. The analytical device used in such situations consisted of a comparison of "commodity funds" (social product in real terms) and "purchasing funds" (personal and government consumption and investment in money terms) which good and up-to-date statistics made possible. The differences between the two were interpreted as excessive money supply. In 1960 the difference was considerable and called for monetary restrictions. In addition, in 1961, during the recession, a monetary reform was undertaken with the purpose of instilling business discipline. Enterprises were forced to increase the share of their own funds in total working capital at their disposal. This share was indeed raised from 7.8 percent in 1960 to 22.4 percent in 1961; the operation was financed out of savings which meant less investment (Vuksanović, 1969). Recession was deepened, retail prices continued to rise at a rate of 6–9 percent, inventories accumulated at a rate of 20–25 percent per year, and monetary authorities decided to tighten up the policy. As a result one enterprise after another found it impossible to settle its debts and mutual indebtedness was expanding at a rate of about 50 percent per year. The Federal government ran out of money and obtained a substantial credit which the National Bank; also the "moneylessness" —a new term coined for the occasion—had to be cured by some credit expansion. All this, of course, ruined the credit balances. The year ended with money supply increased at a rate more than twice as high as the one envisaged. Holjevac complained about the absence of monetary discipline and the fact that the National Bank lost control over credit expansion (1967b, p. 36). But as a consequence the cycle was reversed and the rate of growth accelerated.

The upswing continued through 1963 and with all that excessive (from the point of view of monetary planners) money in the economy, prices were remarkably stable; industrial producer prices rose by one, percent retail prices by four percent. By the end of the that year the upswing developed into a boom, industrial output was expanding at a rate of 15–20 percent per year, and the balance of payments deficit was increasing. Several months later the cycle reached the upper turning point and in the second half of 1964 the recession was already in full swing. All symptoms of the 1961 recession were repeated, and so was the monetary policy. In the second half of 1964 and the beginning of 1965 reserve requirements were raised up to the legal limit of 35 per cent, enterprises were forced to use savings for increases in working capital, investment banks had to use one quarter of their loans for working capital financing, consumer credits were reduced. All this, together with the upheavals caused by the price reform, reduced aggregate demand and output growth from about 15 percent in 1964 to about 4 percent by the end of 1965. Since the tax reform left the Federal government without money, it had to resort to substantial deficit financing, which once again upset monetary planning. But the downswing was arrested for a period of four quarters, and all symptoms normally present at an upwards reversal of the cycle became apparent. However, this time the National Bank had formidable monetary weapons at its disposal and it decided to use them to combat "excessive" liquidity.

For some reason, not explained in the documents, the National Bank established the rule that liquidity reserves of commercial banks held as balances with the National Bank should amount to no more than 6 percent of monetary demand deposits (Perišin, 1969, p. 515). These reserves ran around 10 percent in the second half of 1966. As "purchasing funds" were appreciably higher than "commodity funds"—which is reflected in price increases—it looked obvious that there was too much money in the economy. The National Bank reduced its special credits to commercial banks and put an absolute limit on their credit operations. Consumer credits were further reduced. In 1967 exports were retarded, and so it was decided to depress internal demand even more in order to achieve an export drive. As one might have expected, this did not help exports—in fact their rate of growth was soon reduced below zero in spite of selective export credits—but liquidity reserves were brought down to 5 percent, even lower than planned.

As a result of this anti-inflationary policy output growth was reduced to minus two percent, which had not happened since the Cominform days. The present author estimated losses due to the mistaken monetary policy at eleven percent of social product. Perišin found that gross savings had been reduced from 43 percent of GNP in 1964 to 30 percent in 1967 (1969, p. 517). Unemployment was increasing fast. But price stability was not achieved (see Table 6).

The new system of regulating the money supply proved to be very efficient in reducing money supply to any desired level. This conclusion follows from the foregoing description of its practical operation but can also be illustrated by a series of indices. If we compute ratios of money supply per 100 dinars of transactions, expressed as a sum of gross national product and the output of intermediate goods, we get the following data (Perišin, 1968, p. 63):

TABLE 11

1957	15.3	1963	19.4
1958	15.3	1964	18.3
1959	14.5	1965	14.9
1960	15.1	1966	12.3
1961	15.3	1967	11.3
1962	18.7	1968	13.1

In the three years after 1964 the relative money supply was reduced to 62 percent of its original level. One might be tempted to think that this simply meant increasing the transactions velocity of money. But that is not so; the lack of banking credits was compensated for by involuntary trade credits. The latter amounted to 69 percent of short-term bank credits in 1964, to 138 percent in 1967 and surpassed bank credits almost two times by the middle of 1969.

One other fact is worth noting. Figure 1 shows that boom periods of business cycles occurred in 1960 and 1963 and recession periods in 1961 and 1967. A glance at Table 11 suffices to see the extent to which monetary policy was cyclically synchronized: there was an abundant money supply in the boom and tight money policy in the recession. Consequently monetary policy has been an important destabilizing factor preventing the economy from exploiting its growing potentials.

The second half of 1967 brought the revival and the acceleration of growth continued through 1968 into 1969. Prices were stabilized for a while, inventories reduced, exports soared in 1969 and monetary policy had a relatively easy job to support these favorable trends. It remains to be seen whether monetary authorities—and monetary theorists—have learned the lesson and whether they will be able to avoid making the same mistakes once the trends are reversed.

Public Finance and Fiscal Policy

Budget for a Centrally Planned Economy: In the first two years after the war the new state tried to make the best of the inherited financial system. Taxation was improved in two ways. Before the war a sales tax levied on consumer goods was a major source of government budgetary revenue. That represented a great burden for poorer sections of the population. Next, income tax progression was mild (up to 32 percent) and there were several separate income taxes for various sources of income. Thus people with several sources of income—i.e. the richer ones—could easily evade paying high taxes. It was only natural that the new revolutionary government would make the necessary corrections. The sales tax was reduced from 62.8 percent of government budgetary revenues in 1939/1940 to 46.5 percent in 1946, and separate income taxes were replaced by a single one applied to the entire personal income at increased rates (Finansiski, 1949, p. 25). However, uneven taxation reappeared soon and even in 1969 B. Jelčić complained that the differential tax burden for the same personal income of different taxpayers meant the negation of principles proclaimed and guaranteed by law (1969, p. 159).

When in 1947 central planning was inaugurated, the financial system of the country had to be changed radically. In the old system the government budget used to finance the work of public administration and some social services. That corresponded to the administrative character of the old state. The new socialist state—as described by the Institute of Finance (Finansiski, 1949, p. 16)—acts as an organizer of the entire economy. The targets are annually elaborated in the economic plan and the budget ought to reflect them financially. Each planning organ has its own budget, which is a constituent part of the overall budget. The sum of all financial plans of all ministries, i.e. of all industries, represents an annex to the budget. Thus the budget becomes the financial plan of the entire economy (Finansiski, 1959, p. 11; Matejić, 1958, p. 170). The budget amounted to 64–83 percent of national income (Perić 1964, p. 126). About one half of budgetary revenues was spent on investments.

R. Radovanović describes four principles on which such a budget was based. (1) Centralization of all resources at the disposal of a political–territorial unit (municipality, district, county, province, republic, federation) in the budget of its government. (2) Financing from the budget of all social activities. As far as business firms are concerned, only net revenues are entered into the budget. (3) Concentration of the budgets of all political–territorial units in the Federal budget to ensure central direction in carrying out the most important tasks. This is the famous principle of budgetary monism. (4) As a result of

(3) funds are allocated among various bodies in accordance with their recognized needs and irrespective of their budgetary potentials. Lower bodies are obliged to implement general policy and higher bodies are expected to provide the necessary resources. This had at least one negative consequence. Lower organs were not stimulated to economize with their funds. Instead of trying to expand production in their territories, they were busy in their budget expenditures and exerting pressure on higher bodies to find necessary resources (Radovanović, 1962, p. 1112).

Taxes in such a system are just a technical means for channeling gross profits into the budget (Tišma, 1964, p. 29). The price of a product consists of cost of production, profit and the turnover tax. Profit is generally a small item and is mostly left to the enterprise. If individual planned profit is higher than the average one, three is extra-profit half of which has to be paid into the budget of the higher administrative organ. A planned loss is covered from the higher budget. If achieved profit is higher than planned, half of the difference is left to the enterprise as an incentive. Turnover tax is just a balancing item in an administratively set price. Since it is charged on all commodities and is paid as soon as a commodity is shipped, it is also used as an indicator of how the implementation of the plan is proceeding.

In order to accommodate productivity change in such a rigid price structure the "decrease in full cost of production" was explicitly planned as a separate item. This decrease is partly paid into the budget and so a rather unusual new type of tax was created. Finally, various types of prices, discussed in the section on administratively set prices generated so-called commercial profit, which was mostly absorbed by the budget.

In 1949 the four items enumerated were (in billions of dinars): turnover tax 66.6, share of profits 4.6, decrease in cost of production 3.8 and share of commercial profits 13.1 (Tišma, 1964, p. 96). Turnover tax represented, of course, the bulk of budgetary revenues.

The major proportion of budget revenues came from the business sector. Taxes paid by the population were steadily decreasing in importance, from 22.4 percent of all revenues in 1946 to 9.7 percent in 1952. As a consequence taxation of the population was governed by extra-fiscal considerations. In 1950 the tax on income earned in the state sector was abolished. (It was to be reintroduced only in 1960). This did not matter much, since wage and salary differentials were greatly reduced and income distribution was extremely egalitarian. But income taxes were retained for the private sector and the progression was rather stiff. For peasants the tax rates went up to 70 percent in 1947 and up to 90 percent in 1948, as compared with the flat rate of 3 percent for the members of peasant work cooperatives (cooperatives organized similarly to state firms) (Finansiski, 1949, p. 34). This tax policy was inspired by the idea of the class struggle and was aimed at inducing peasants to join cooperatives.

The policy of stiff taxation of peasants and artisans was continued also later and for the same reasons. In agriculture it was discontinued after the second agrarian reform in 1953, which reduced the maximum size of agricultural estates to 25 acres and so eliminated any possibility of capitalist development. In 1954 taxation on the basis of cadastral[11] income was introduced, and rates were lowered. Both proved to be stimulating. It is interesting to note that Radovanović described the tax on cadastral revenue as an instrument designed to replace compulsory deliveries while making sure that a minimum output will be produced (Hanžeković, 1967a, p. 91). There was no possibility of capitalist development in handicrafts either, because artisans could employ at most five workers. However, public opinion held that there was something vicious about private business. Tax rates were substantially reduced only in 1963 (Hanžeković, 1967b, p. 33). The policy of containment continued until the ownership discussions in 1967 analyzed in the section on Ownership Controversy. In the meantime the number of artisan shops was substantially reduced, which caused economic difficulties.

---

[11] Cadastral revenue is the value of the average yield of a specific land category under average weather conditions and using an average land cultivation technique.

Taxation Experiments: After the French Revolution in 1789,—remarked J. Lovčević, the Constituent Assemly abandoned taxes in favor of contributions. After the Yugoslav Revolution a law on taxes passed in 1946 declared that a tax was "a contribution. . . given to the state for economic development, cultural advancement . . . and for the maintenance of the state apparatus" (Milatović, 1967, p. 34). In spite of all its protests of public finance experts,[12] the term stuck. From 1952 enterprises have been paying contributions (turnover tax representing an exception) and individuals taxes. Contributions somehow emanated from social property, taxes from private property. Since the 1965 tax reform contributions have become synonyms for direct taxes or taxes levied on labor income and the term tax is used to denote various forms of turnover tax or property tax. The terminological confusion did not matter very much. But lack of professional competence in designing an appropriate taxation system did matter. In the period 1952–1965 the tax system was changed five times with obvious consequences as far as the efficiency of conducting business was concerned.

In 1952–1953 the system of AF rates—whose rationale was discussed in the section on Distribution Policy—predetermined the taxation system. Out of accumulation and funds obtained by the application of a rate, prescribed by the social plan, to the net product of an enterprise, the social contribution was paid to the budget. It contained social insurance payments, was proportional to the wage bill and was paid at the flat rate of 45 percent. Wage bills above the standard prescribed were taxed at steeply increasing rates. A tax on extra profits was envisaged by law, but never applied due to technical difficulties (Tišma, 1964, p. 97). Turnover tax was greatly reduced, and amounted to 9–14 percent of budgetary revenues (Jelčić, 1967b, p. 14). Its task was to absorb monopoly profit and to influence price formation (Radovanović, 1953, p. 62).

The system of AF rates helped to eliminate administrative ties between enterprises and planning authorities, but soon degenerated into administrative determination of AF rates for each individual enterprise. It had to be replaced by a system working more in a market fashion. It was not clear how to design such a system. It seemed advisable to make use of the experience of traditional market economies. Instead of net product, profit was the base of taxation for the next four years (1954–1957). Wages became part of costs of production. Profit was taxed at a flat 50 percent rate. The other half of gross profit was used for contributions to SIF's for supplements to basic wages, for enterprise funds and for some other purposes. Wages from profits were linked with contributions to local budgets which amounted to same sort of progressive payroll taxation. A tax on monopoly profit was envisaged but never applied because it proved impossible to established which part of the income resulted exclusively from the work of the collective. The share of the enterprise (wages and undistributed profits) gradually increased to one third of net product generated (net product included turnover tax) (Tišma, 1964, p. 99).

---

[12] Fiscal theory distinguishes taxes, contributions and stamp duties. A tax is a compulsory payment for, in principle, no specific service. A contribution represents a compulsory payment for a specific service and in principle covers the cost. Stamp duty is a payment for a specific service at the initiative of the payer, but it bears no necessary relation to the cost.

In this period two interesting new taxes were introduced. Mines, hydroelectric power stations and some other firms were to pay rent. Artisans and peasants were obliged to pay tax on hired labor. The latter tax was insignificant in quantitative terms, because only one-eighth of the artisans and almost no one among peasants hired labor, but served as a reminder that hiring labor meant exploitation.

Wages as part of cost of production were deemed inappropriate for a self-management system. Thus the new system, inaugurated in 1958, was based on the distribution of total enterprise income. That was a switch back from profit to net product, reduced for turnover tax and some other items. There was also a terminological change: wages and salaries were replaced by personal income. With many changes the system lasted until 1964.

The main tax, surpassed only after 1961 by the turnover tax, was the contribution from income. The rates were progressive up to 80 percent. Tax progression was in 1961 replaced by a flat rate of 15 percent and a surtax of 25 percent. In the meantime another development took place. It appeared reasonable to link collective consumption and public services to the level of personal incomes earned in any particular territory. For this purpose contributions to budgets were made out of the wage bill. In 1958 these contributions were progressive, in 1959 a flat rate of 11 percent was charged; the rate was increased to 15 percent in 1963. In 1964 some tax rates were reduced, and mining taxes and contributions to SIF abolished. The abolition of progressive rates led to a reintroduction of the progressive personal income tax (Tišma, 1964, p. 207). This indicated that the economic functions of the payroll tax and personal income tax, as discussed below, had been confused. The share of the enterprise in its net product increased to about one half.

In order to increase this share still further, the last reform of 1965 abolished all contributions from enterprise income. The share in net product jumped to about two-thirds. Since then enterprise taxation has rested exclusively on payroll taxes. If we count social insurance contributions, labor has been made about 60 percent more expensive than necessary. This has serious consequences. Before 1960 taxation created capital saving inducements (Pejovich, 1964): in a labor surplus economy that was rational. Since 1964 taxes have stimulated labor saving practices. Enterprises did in fact react: coal was being replaced by oil, cotton growing and cattle raising by wheat cultivation and so on, and thousands of workers became redundant. Further, flat rates introduced an awkward rigidity and tended to intensify cycles. Finally, the abolition of progressive payroll taxation after 1958 and the lifting of wages control in 1961 meant that two important checks on inflationary pressures were eliminated. We have already discussed the consequences.

Taxation experiments have clearly not been completed. Is there anything one could say about how an appropriate taxation system ought to be designed? On various occasions the IES has made suggestions in this regard, and they may be summed up as follows. The equilization of personal income distribution can be achieved by the familiar progressive personal income tax. There is no need to tax profits, even less to tax them progressively, since capital is socially owned. But there is a need to tax payrolls and to tax them progressively. In order to do this wages ought to be standardized by applying accounting wages for certain categories of skill. (The skill rating should, of course, not be left to enterprises themselves, just as school diplomas are not issued by pupils themselves.) When faced with the alternative of either losing a greater part of the "excess wage fund" through taxes or using that money for development purposes the working collective will often opt for the latter. This will check wage increases in the most profitable enterprises—which have continually been generating wage pushes—and expand their investment, increasing the supply of their products relative to demand and lowering prices. Labor should be made as cheap as possible (for the enterprise, of course, not for the workers) in order to stimulate labor intensive production. If some taxes still prove necessary, they may be levied on the enterprise income at a flat rate. Such "contributions from income" may be considered as a self–management counterpart to the familiar value–added tax.

While direct taxes have received little attention in professional economic literature, turnover tax has been extensively discussed. And with good reason. It survived through all tax reforms as one of the principal taxes. Since 1954 the share of turnover tax in total budget revenues has oscillated between 29 and 43 percent (Hanžeković, 1967a, p. 28). By 1964 six kinds of turnover tax were in operation (Lazarević, 1965). Producers' turnover tax was inherited from the days of central planning It was levied on some 250 products at rates varying between 2 and 81 percent: it was contained in producers' prices, represented a part of enterprises' gross receipts and was collected at the time the invoice was issued. It was easily and quickly collected, even before bills were paid, and was liked by the government. It was also used as an instrument of price policy. In order to provide independent sources for communal budgets, in 1956 a communal sales tax was introduced. In 1961 owing to the abolition of progression in enterprise income taxation the government ran short of money and introduced the one percent general turnover tax. This was a multiple–stage tax and was intended to reduce the number of middlemen between producers and final consumers: however, apparently no effect of this kind was achieved (Hanžeković, 1967b, p. 47). There was then also purchase tax on specific products, service sales tax and duty on real estate and other transfers.

Producers' turnover tax has been severely criticized. Both the government and the enterprises tended to abuse it as a price formation device. In twelve years its tariff was changed almost one hundred times (Jelčić, 1967c, p. 4). Its handling required a large amount of working capital on the part of the enterprise. It tended to distort prices, and so did the multiple stage general turnover tax. In the case of exports, tax deductions had to be computed and made. For all these reasons the two kinds of turnover tax were abolished and in 1965 replaced by a sales tax levied on consumer goods in retail trade, added to retail prices, charged directly to buyers and collected when the commodity was sold. But a retail trade sales tax cannot be changed often and cannot be differentiated for many products. Thus its use as a price formation instrument is rather limited. It is now primarily a device for collecting budget revenue.

Budget for a Self–Government Economy: A budget is more consistent with a centrally planned economy the more all–embracing it is. Ideally all financial transactions of the economy are to be regulated by the budget. It is the other way round in a self–government economy. Here the budget ought to be restricted to as small a section of the economy as possible in order not to interfere with the economic activities of work collectives. Ideally the budget should cover only the activities of various state agencies. In this respect the 1952 reform initiated three important developments. They were related to the organization of the non-market sector of the economy, to the creation of various social funds and to the decentralization of budgetary revenues and expenditures.

The Yugoslav tradition had made a sharp division between enterprises (poduzeća) and institutions (ustanove). The former were business establishments, the latter were financed from the budget and roughly corresponded to non-profit institutions in the USA and elsewhere. Since the latter depended on the budget, i.e. on the government administration, for their revenues, it was clear that self-management had little chance of developing. Thus institutions that performed public services and could be financed partly or wholly by selling their services[13] were separated in a special group of "institutions with independent finance." Gradually it became evident that there were two fundamentally different types of public services: the one (government administration, judiciary, police, defense) rendering various administrative services to society, the other (education, science, medical care, etc.) increasing the welfare of the members of society. It seemed appropriate to finance the former from the budget ("public expenditure") and to organize them in a more or less traditional fashion, but the latter ("collective consumption") required a different approach. M. Hanžeković suggested that taxes be used to finance the former and contributions the latter (Hanžeković, 1967a, p. 17).

Next, while there was to be a free market for the short-run operations of enterprises, it appeared advisable to retain substantial central control in the field of capital formation. But capital financing was to be on a credit basis and budgetary financing implied grants without repayments. Thus investment resources were separated from the budget and concentrated in investment loan funds. The budget continued to finance investment projects in the nonmarket sector (schools, hospitals, etc.).

In 1952 social insurance had also been separated from the budget. This decision was motivated by the fact that social insurance could be efficiently operated as an independent social service *under a social self-government regime.* The latter meant that the governing bodies were composed of representatives of various social interests (physicians, social workers, citizens, government representatives).

Very soon there was a proliferation of various funds for housing, for advancement of agriculture and forestry, for roads, for cultural activities, for education etc. Many of these funds had their independent management bodies and obtained their resources from special contributions or from budgets. Hanžeković suggested the following three-fold classification (1967a, p. 13): (1) funds for capital formation (SIF) or for financing public services; (2) funds for financing without repayment obligation or for granting credit; (3) with self-government bodies or without. Definite trends have appeared in further developments. Loanable funds were mostly transferred to the banks. Funds without independent management bodies are used as often temporary instruments of budgetary financing for special purposes. The third category, permanent funds with independent self-government, represents an innovation.

The social-insurance fund set an example. A decade later the example was followed by education. At first, T. Konevski remarked, that was just a transmission mechanism in budgetary financing (1968, p. 163). But in 1967 Education Unions were formed to operate the funds. Assemblies at communal and republican levels vote money to be allocated to education funds. Education Unions—self-government bodies composed of representatives of schools, outstanding figures in cultural life, government agencies—distribute the money by negotiating the services to be rendered by various educational establishments. In 1969 Research Unions were formed. They operate funds for research work created in 1960. Unions are shown in the quasi-market sector of Figure 2.

---

[13] The law of 1959 changed this condition into "institutions organized according to the principles of social self-government." The institutions were renamed "independent institutions." In 1965 they obtained the status of work organizations with the same self-management rights as enterprises.

Hanžeković points out that in 1965 funds absorbed 8.8 percent and institutions 14.2 percent of national income, which had to be compared with total budgetary expenditures that amounted to 20.1 percent of national income (1967a, p. 14). Institutions obtain about one third of their income from selling their services to direct buyers (to the market), 50–60 percent of their revenues come from various funds (quasi-market) and only one-tenth derives from budetary subsidies. Such a structure of revenues enabled the non-market (not non-profit, because they do make profits) institutions to gain a considerable amount of independence. Also, they established closer contacts with the buyers of their services and with the rest of the economy. Is there, one might ask, any economic activity in which an Archeology Department of a University, a museum or art gallery can engage? Yes, there is, though perhaps not directly. Tourist agencies and hotels may be, and in fact are, interested in financing the development of an archeological site, a local museum or art gallery. Sometimes these are rather roundabout ways for achieving certain goals, but if they eliminate government control and increase independence, the price may not be too high. Yet there are other costs involved. Konevski points out some of them (1968, pp. 128–65). To administer a fund an administrative apparatus has to be set up. Unlike business enterprises in the market, a school or a hospital is in an inferior position when it negotiates contracts with the funds. Commercialism may and does have detrimental effects in such fields as culture, education, science or medical care. The consumer may be, and often is, victimized. Since it is too early to evaluate the working of the system, one can only invoke the wisdom of the ancient Greeks concerning the organization of human affairs: right proportions, no extremes.

The creation of funds and the establishment of self-financed institutions represent two aspects of decentralization. As a consequence the share of budgetary revenues in national income was reduced from one-third in 1952 to one-fifth in 1967. The third aspect of decentralization was related to the division of revenues among budgets of various socio-political units. The federation was gradually transferring its responsibilities for various social services to republics and communes. As a result the share of federal expenditures in total budgetary expenditures dropped from 74 percent in 1952 to 53 percent in 1968. The trends have been reversed as compared with what happens elsewhere.[14]

The division of budget revenues among various budgets is a somewhat complicated technical problem. Not less than five laws in the period 1952–1965 tried to solve it and with only limited success. In theory there are two possibilities: a separation of revenues, and joint revenues. Both have been tried out at one time or another.

---

[14] In the USA the share of federal revenues in total budgetary revenues increased from 42 percent in 1890 to 75 percent in 1954; in Switzerland federal expenditures amounted to one half of cantonal expenditures in 1913 and to 111 percent in 1958 (Bogoev, 1964, p. 10).

After 1952 the budget monism of a centrally planned economy was replaced by a budget pluralism better suited to a self-government economy. The former budgetary system was based on participation in joint revenues, the higher governmental bodies determining the conditions of participation. If lower budgetary units were to be made more independent in the development of revenue sources in their own territories, a system based on a separation of revenue sources seemed more appropriate. Thus sources of revenue were allocated to budgets at various levels. Only the federation was entitled to introduce new taxes, but, if introduced, taxes had to be immediately allocated to specific budgets or funds. In principle every unit was to cover expenditures from its own revenues. This principle was not fully implemented, but there was a great change as compared with the former practice. In two characteristic years republics and communes obtained their revenues in the following ways (Radovanović, 1956a, p. 445):

TABLE 12

	1948	1954
Own revenues	53.7%	72.5%
Participation in joint revenues	43.3%	22.5%
Federal subsidies	3.0%	5.0%

With many changes this system lasted for nine years (1952–1959). Its main shortcomings, as described by Radovanović (1962, p. 115) and K. Bogoev (1964, pp. 188–90) were two. Sources allocated to lower units were not sufficient to meet the recognized needs. Deficits were substantial and were covered by sharing in revenues and by subsidies. These were discussed every year anew, which made lower units very dependent on higher authorities. Next, the lack of objective allocation criteria generated a bargaining process. For both reasons the system failed to provide stability and incentives.

In the period 1960–1964 the budgetary system was again based on participation. Separate sources were allocated only to the federation (they covered 90 percent of its revenues) and to communes about 20 percent of their revenues). Republics and districts had no separate sources. The participation of all units was determined by federal and republican laws. The higher units could not arbitrarily select more favorable sources for themselves. In order to eliminate another source of arbitrariness, participation rates were not differentiated according to sources as before, but instead one single participation rate was applied to all sources of revenue. Participation rates were increased for less developed units, and if this was not sufficient, subsidies were granted. Increased shares and subsidies were to be determined on the basis of the funds needed for carrying out "mandatory tasks and services." However, since objective criteria were not established, the familiar arbitrariness crept into the process. In 1960 only 9 percent, and in the following year only 3.6 percent of all communes were able to cover their needs in the regular way (Bogoev, 1964, p. 205). About one-half of all communes had to rely on both increased participation shares and subsidies. What was intended to be a corrective device turned out to be the main instrument for balancing budgets of lower units.

The 1965 tax reform introduced the separation principle once again. The sources were allocated as follows. Taxes on personal income and sales taxes may be introduced by all socio-political communities. Apart from that, taxes on property (and some other taxes) belonged to communes, estate duties to republics and customs duties to the federation. Communes and republics are empowered to decide independently what kinds of revenue to introduce for their territories and to fix the tax rates. There are two safeguards. The federal government can fix temporarily the limits for the tax rates set by republics and communes. Communes and republics are legally obliged to cooperate with one another in fixing the level of their revenues in order to assure citizens equal treatment. Republics and provinces are entitled to federal subsidies provided their per capita revenue is below the Yugoslav average and they have exhausted all possibilities for collecting revenue through taxation of personal income, in conformity with the economic potential of their population (Turčinović, 1968).

This time the criterion for subsidies has been defined somewhat more precisely. But it has also been criticized. Hanžeković argues that approximately equal budgetary revenue per capita cannot be an appropriate criterion. Instead appropriately defined necessary and justified expenditure should provide a basis for allocations (1967a, p. 7). In fact this seems to be *the* problem of the Yugoslav budget system. Yugoslav territories are extremely unevenly developed. Per capita income in the Republic of Solvenia is 5.4 times higher than in the Autonomous Province of Kosovo. Communal budgetary revenues are, of course, even more unequal: in 1965 the most developed commune in Slovenia obtained per capita revenue almost 16 times higher than the least developed commune in Kosovo. Such extreme differences inevitably ruined all schemes in which allocation criteria were not precisely defined. Konevski complains that in the new system more than one-half of communes in Serbia have to rely on subsidies, which is inconsistent with the philosophy of self-government (1968, p. 116).

In 1968 the government asked a research institute to study the problem. A group under the chairmanship of P. Sicherl prepared a voluminous report (Sicherl *et al.*, 1968). Sicherl finds that although differences between the developed and the underdeveloped regions in per capita income are extreme, differences in nonagricultural income per worker are small. He used a a special statistical method developed by his colleague B. Ivanović (1964) to establish that the distance between developed and underdeveloped regions is appreciably greater in the economic sphere than in the sphere of social services and living standard. In a later article Sicherl argues that it is easier to reduce the distance in the latter sphere (in terms of flows of services) than in per capita national income (1969). As a basis for subsidy computations, Sicherl takes accounting budgetary revenue which he defines as revenue obtained by applying the average Yugoslav tax rates to actual tax sources in the region. The dilemma of whether policy should be based on the equalization of needs or of revenues is resolved in favor of revenues, on the ground that it is difficult to determine needs in an objective way and that to do so is also inconsistent with the philosophy of decentralized decision making. There follows a long and involved discussion of the most appropriate method of determining standard revenue. The difference between the standard and the accounting revenue is to be covered by federal subsidy. Sicherl's Report has been discussed in government and parliamentary committees but has not produced practical results as yet.

Communal Economy: In daily life every man appears in a double capacity: as a producer and as a citizen. Thus direct democracy will also have two aspects: one relating to the work place, the other to the territory where citizens live. As members of working collectives, people engage in self-management. As inhabitants of towns and villages, they manage their affairs by establishing local self-government. The territorial association that corresponds to the collective at the work place is the commune.

There has been a strong tradition in local government in Yugoslavia since the days of the National Liberation War. People's Liberation Committees, as local government bodies, worked with great independence, initiative and resourcefulness to supply the partisan army and organize daily life in the liberated territories. It is hardly a matter of chance that the first People's Committee and the first Committee of Workers' management appeared simultaneously in the fall of 1941 in the mining town Krupanj. People's Committees continued to exist after the war, but than as components of a rigidly centralized system. The system was based on the principle of democratic centralism, which meant that higher bodies could abrogate decisions of People's Committees.

This practice was radically changed in the fateful year of 1952. The principle of democratic centralism was replaced by the principle of legality control (Dordević, 1957, p. 24). District People's Committees became organs of self-government and Communal People's Committees organs of local government. District Committees had assemblies with two houses: one composed of political representatives, the other of representatives of producers. The next crucial step was taken three years later. The 1955 law on local self-government proclaimed that the Commune was "the basic political-territorial organization of self-government by the working people and the basic socio-economic community of the population on their territory." The Constitution of 1963 changed the phrasing slightly to make the commune "the basic socio-political community." The development of the communal system has been greatly influenced by the historical example set in 1871 by the Paris Commune, "that finally discovered political form in which emancipation of labor can be carried out" (Marx). It is useful to notice, as D. Milivojević points out, that the commune has not been conceived as just a form of otherwise familiar local government. It is a community of those living, working and producing, satisfying their basic needs, and realizing their civil and self-governing rights in a particular territory (1965, p. 8). For a while districts retained certain coordinating functions and then gradually withered away.

Since the commune is a territorial association, one of the first problems to be solved was to determine the size of the territory. The problem was solved by practical experimentation over the period of a decade. Consistent with central planning was a hierarchy of governmental levels. There were three levels below the level of republic: county (oblast), district (kotar) and local committee (mjesni narodni odbor). In 1951 counties disappeared. The orientation towards a market economy made excessive administrative fragmentation—there were more than 7,000 local committees—unnecessary and so in 1952 the number of local committees was halved and committees were replaced by communes. In order to bring local government closer to citizens, in 1955 the commune was made the basic self-government unit. Since, however, the commune was expected to exercise a wide variety of functions, its territory had to be increased. Table 13 depicts the process of territorial transformation. Each new law on territorial changes, remarked E. Pusić, was announced as the last and the definite one (1968, p. 245).

Communal territory was growing larger and larger and by 1967 the average poulation size of the commune (40,000 in 1967) almost reached the poulation size of the district at the beginning of the process (48,000 in 1952). The district became superfluous and disappeared. The larger commune was more efficient, but less self-governing; that is why the new Constitution provided for the creation of local

TABLE 13.—NUMBER OF TERRITORIAL UNITS OF LOCAL GOVERNMENT (END OF THE YEAR)

	1948	1952	1955	1967
Local committees/communities	7967	—	—	4968*
Communes	—	4052	1479	501
Districts	427	351	107	—

* 1965.

Sources: *Jugoslavija 1945–1954*, pp. 35–36. *SGJ–1968*, p. 62. *Yugoslav Survey*, 1965, p. 3296.

communities. These were to be self-governing communities of citizens in rural and urban localities concerned with all activities connected with the satisfaction of the needs of citizens and their families. J. Duričić describes three functions of a local community: it is (a) a form of self-government including traditional political activities, (b) a unit of town planning and (c) an organization taking care of some social services, public utilities, etc. (1965). Pusić is rather skeptical about local communities contributing in any important way to self-government. In his view their activities are too restricted to be particularly attractive to the citizens and in a modern urban setting territorial closeness per se generates no specially active social ties (1968, p. 243). There are 27,706 localities in Yugoslavia, and by 1965 statutes of communes provided for the creation of 4,968 local communities (7.7 percent of communes did not establish local communities at that time). The organizational circle seems to have been closed: communes have replaced districts and local communities have replaced local committees. But considerable social experience has been accumulated in the process.

Apart from exercizing the functions of traditional local government, which include local politics, public utilities, education, social welfare, etc., a commune is also responsible for other aspects of local life. D. Miljković explains this in detail. The commune is expected to harmonize individual and social interests. It is responsible for social property, either under its own control or "belonging" to enterprises. It takes care of economic development and cultural advancement. It coordinates all economic, social and political activities on its territory, prepares a social plan and makes it possible for citizens to participate in the process of social decision-making (Miljković, 1961: Jelčić, 1969). But communal self-government is a contradictory institution, remarked Djordjević, as it carries with it forces of unification and disintegration Both forces will soon make themselves felt.

The 1955 law was preceded by extensive discussions about the functions of the commune. In a paper presented at the annual meeting of Serbian economists in 1954 J. Davičo maintained, and those present agreed, that a labor managed enterprise had no incentive to embark upon substantial capital formation. In his opinion large investment would imply creating a new enterprise which would be equally labor managed and so could not be dominated. For this reason Davičo argued that the commune was "the natural investor in our circumstances" (1954, p. 192). As Table 10 shows, communes indeed became large investors. In 1964, when a maximum was reached, 25 percent of all investment in fixed capital was financed by communes (and districts). Since 1959 communes have been entitled to initiate the setting up of all kinds of enterprises, to bring about mergers or carry out liquidations (Bogoev, 1964, p. 129). However, the last economic reform put an almost exclusive reliance on enterprises as far as capital formation was concerned, and by 1968 the communal share in investments dwindled to four percent. But this left other economic functions of the commune intact. In cases of failure of an enterprise, the commune shares a good deal of the financial responsibility involved. The commune also gives guarantees for credits and loans granted by the banks to enterprises located on its territory.

For people accustomed to central planning, i.e. to administrative methods in running an economy, it was difficult to imagine a really free market. They were determined to get rid of governmental controls. It seemed obvious that the best way to achieve that was to replace it by communal control. The self-governing commune would tell enterprises what to do and how to behave. In 1954 and 1955 communes were empowered to determine the needs of enterprises and to distribute their profits after federal taxation. Since they were entitled to determine their shares in profits and since they were independent in budget expenditures, communes taxed incomes of enterprises more than the latter could bear. The consequence was a general price rise as shown in Table 6 and Figure 1. In 1956 taxation rights of the communes were again regulated by federal laws (Radovanović, 1956b, pp. 113–16; Bogoev, 1964, p. 166).

Gradually romantic views of conflictless communities, local or otherwise, had to be revised. Hopes have been directed towards an impersonal market mechanism, but expectations have again been a little unwarranted, I am sorry to say as an economist. But at least people were willing to learn from experience. Enterprises gained communal boundaries. Communal banks, which kept appearing in the period 1948–1964, became just commercial banks. The approach to communal economy, self-government and life became far more sophisticated. The actual economic, social and political importance of communes has not decreased, though lately republics show a tendency to encroach upon communal finance.

In an excellent study Bogoev surveys the development of communal finance (1964). In this context one difficult fiscal problem—adequate finance for administrative and, in particular, for social services—may be singled out for closer scrutiny. Bogoev and Petrović point out that the 1957 Resolution of the Federal Assembly on public expenditure and collective consumption which together comprise "general consumption" in Yugoslav terminology as distinct from privately financed consumption) demanded that such expenditure be tied to the economic potentials of the area in question (Bogoev, 1964, p. 179; Petrović, 1968, p. 57). Later the new constitution insisted on the principle of work performed as one of the taxation criteria to be applied to revenues of socio-political units. Tax laws interpreted these two principles to mean that taxes should be collected in proportion to personal income. For this reason the proportional payroll tax gained in importance until after 1964 it became the only tax paid by the enterprises. Since collective consumption is a kind of personal consumption collectively financed, it seemed just and proper to link it with personal incomes earned in a particular territory. The payroll tax was made even more attractive when it was arranged that it be paid into the budget of the commune where people lived and not where they worked or where the enterprise head office was located. It is only recently that the short-comings of the payroll tax and the fallacy in the reasoning by which it was introduced have begun to be discussed.

Let me close this section by a brief review of the main activities of a commune. What communes do is best seen from a breakdown of budgetary expenditures, as shown in Table 14.

Public utilities, education, infrastructural investment and public administration are activities controlled by the commune more than by either republics or federation. Bogoev points out that the communal share in total budgetary expenditures is one of the highest in the world (29–35 percent or 50 percent without defense in Yugoslavia as against 30 percent in Western Germany, 25 percent in Switzerland, 22 percent in Austria and 20 percent or 35 percent without defense in the USA) (1964, p. 329). Whether this share has reached the upper limit remains to be seen.

Fiscal Policy: I add this section for the sake of completeness. But it might as well have been omitted. Strange as it may sound, there is no fiscal policy in Yugoslavia. In fact, this is quite consistent with the belief in the absence—or with the ignorance of the presence—of business cycles.

Fiscal policy can affect aggregate demand via the revenue or the expenditure side of the budget. The revenue side, taxation, has been recognized as a legitimate tool of fiscal policy in theory and is sometimes used in practice. Producers' turnover tax has been occasionally used to affect the general level of prices in order to absorb excessive purchasing power. Otherwise numerous tax changes have been made in order to affect individual prices or to increase the discretionary power of enterprises over their incomes and have not been intended to affect aggregate demand. To a certain extent selective turnover tax reductions have occasionally had price stabilization effects.

The federal government occasionally ran a substantial deficit in recession years, as for instance in 1962 and 1965. But that was purely accidental, a consequence of the combined effects of tax reforms and the lack of revenues. Textbooks on public finance, written invariably by people with training in law, keep on reminding students of the time-honored principle of sound finance: the balanced budget. And since governments on all levels were not too scrupulous in their spending practices, insisting on balancing the budget was quite justified. Bogoev points out that the budget has always been balanced when presented to the Federal Assembly for acceptance and that only in implementation would deficits appear. Deficits have amounted to 10–15 percent of the federal budget and up to 5 percent of republican and communal budgets, but have been much larger for extrabudgetary expenditures (investment, social insurance) (Bogoev, 1966, p. 159).

TABLE 14.—BUDGET EXPENDITURE IN 1966

	Total expenditure	Federation	Republics and provinces	Communes
Total expenditure effected	100	45.8	19.4	34.8
Education	100	0.1	21.5	78.4
Science and culture	100	5.3	58.1	36.6
Social welfare and medical care	100	52.0	11.6	36.4
Public utilities	100	—	16.2	83.8
Public administration	100	16.7	40.0	43.3
National defense	100	99.7	—	0.3
Infrastructural investment	100	5.3	38.8	55.9

Source: Turčinović, 1968, p. 71.

The first public debate about fiscal policy took place in 1967. At an economic conference in Ljubljana Bogoev (1967), Hanžeković (1967b) and Jelčić discussed the absence of fiscal policy in Yugoslavia and made various suggestions. Bogoev quotes the Resolution of the Federal Assembly on Economic Policy in 1967 which stated that there was excess demand and that not only had all budgets to be balanced but also reserves had to be accumulated. As our Figure 1 shows, Yugoslavia experienced an unusual depression in 1967. Bogoev also points out that proportional tax rates levied on payrolls have cycle-intensifying effects and that the small amount of transfer expenditures (unemployment compensation, debt repayment subsidies) limits the possibilities of an effective anticyclical policy. In the post-war period the federal government raised three internal loans (for the First Five Year Plan, to counteract the effects of the Cominform economic boycott and to finance the rebuilding of Skopje, destroyed by an earthquake). The sole purpose of these loans was to transform a part of personal consumption into investment. Bogoev believes that the rigidity of the existing fiscal pluralism may be softened and an effective anticyclical use made of appropriately designed federal budgetary subsidies to other budgets.

B. Šoškić is the only other economist who has made written contributions related to fiscal policy (1969a). Šoškić was primarily interested in the expansionary effects of public works. In his view the most appropriate objects of increased public financing are: housing and communal construction, road construction, land reclamation and irrigation projects, and power generation projects. Such investment projects are desirable also because of their very low import content, as was pointed out by the IES. Šoškić added that they were also very labor intensive, which is of great importance for a labor surplus economy (1969b).

VI. *Self-Government, Market and Socialism*

Limitations of space preclude discussion of two important lines of economic policy, agricultural policy and regional development policy. But there is one permanent theme of Yugoslav social science discussion which cannot be neglected: the interrelationship between socialism self-government and market. Recent discussions of this problem will be surveyed in this concluding chapter.

I have already discussed the familiar contention that socialism and markets ("commodity production") are incompatible. It was the basis of P. Sweezy's criticism of Yugoslav economic policy as a "gradual transition from socialism to capitalism" (1964). Sweezy argues that the market restricts socialist relations and transforms social ownership into a sort of collective ownership. Material incentives and market orientation necessarily generate a profiteering mentality. The evaluation of social usefulness by profit is characteristic of a capitalist system. Gadgetry and acquisitiveness replace socialist values. This sort of criticism is fairly common. J. Djordjević argues in reply that the undesirable social phenomena are the result of industrial civilization and not only the consequence of the market. The abolition of the market means a return to étatism and state property. Self-government implies free disposal of earned income and, more generally, business autonomy which, in turn, implies markets. If this is not understood, the alternative is an old one: the eschatological idea of state rule and the re-education of man. "Man would be placed under the tutelage of the state (or party, or some other mechanism) to be prepared and educated, so that one day he may become an adult socialist subject" (1966, p. 96).

Yugoslav economists are quite unanimous in believing that the market ought to be maximally exploited as a device of economic organization. Philosophers, however, have their doubts. M. Marković, a leading philosopher actively interested in economic affairs, believes that initial forms of workers' self-management cannot be achieved without material incentives which imply market competition. However, if exclusive reliance on money relations became a permanent feature of the society, self-management might gradually degenerate into a sort of capitalist cooperative. If the results of work were permanently evaluated in terms of income, and if the desire to earn as much money as possible became a permanent and basic interest of a worker, this would produce a personality type not basically different from the type produced by a capitalist society (1965, p. 70).

Referring to Marx, some of my philosopher colleagues declared that socialist commodity production was a *contradictio in adjecto*. In Marx's sense commodity production implies market relationships which result in "commodity fetishism" and various alienation phenomena. I tried to clarify matters in the following way. The familiar statement that commodity production generates capitalism ought to be reversed. Commodity production existed in slavery, feudalism, and capitalism as well as in étatism. It clearly did not determine all these socio-economic systems; on the contrary, it was determined by some more fundamental social relationships and was shaped by respective social systems. Thus, for instance, capitalism resulted from private ownership, étatism from state ownership. Since there are so many types of commodity production, it need not be surprising if we also find socialist commodity production. The elimination of private ownership does not necessarily produce socialism, although it may restrict the role of the market considerably. If private ownership is replaced by state ownership, capitalism is replaced by étatism and commodity fetishism by office fetishism. In both cases relations among people are reified, social inequality preserved, class exploitation continued, essentially human existence made impossible. In socialism social ownership makes social capital equally accessible to anybody while the authoritarianism of a privately managed or a state managed firm is replaced by self-management. In this context the market and planning are not goals but means. If a working collective is to be really autonomous in economic decision-making, the market is indispensable. But planning contradicts the business autonomy of an enterprise and so the choice is between planning and the market—says a time-honored fallacy. In fact social planning, far from restricting, enlarges the autonomy of enterprises for at least three reasons: (1) it reduces uncertainty which is the basic restriction on free decision-making; (2) it increases the rate of growth, the market expands and so the number of available alternatives increases; (3) it equalizes business conditions and so makes the success of a producer less dependent on external conditions which he cannot control and which are economically and socially irrational (Horvat, 1968c).

The nature of the relationship between the market and the plan is a frequently discussed subject. Plan and market have been traditionally contrasted as two separate mechanisms. But some economists try to develop a monistic approach. Bakarić argues that there can be no contrasting, that the law of value reigns supreme and that planning is just one, although the most important, element in it (1963, p. 52). This statement seems to be the reverse of what I said in the preceding paragraph and in the section on decentralization, but the contradiction is more apparent than real. What Bakarić tries to do is to combat the voluntarism of étatist planning and to show that there is an objectively given framework within which planners are obliged to move. Maksimović understood this statement to mean too much *laissez-faire* to his taste. He criticizes the inconsistencies of the officially proclaimed economic policy and warns that an insufficiently controlled market causes damage to individuals (negation of distribution according to work), and to enterprises (different business conditions in various industries) as well as to the society at large (less than optimal production). All this tends to generate an ideology which maintains that socialism is not economically superior to organized capitalism, that inequality and exploitation are products of human nature and cannot be eliminated (1964).

D. Mišić sees the shortcomings of self-management, as it exists today in Yugoslavia, primarily in the fact that it is confined to the enterprise. Investment resources are not allocated rationally; in the present situation self-management and planning contradict each other, the socialist distribution principle is negated and there is a tendency for group ownership to arise. As a result a *laissez-faire* approach is extolled. Mišić suggests that the self-management structure be completed upwards. He believes that the integration processes, which was discussed in the section on enterprise, are neither fast enough nor quite appropriate. Mišić pleads for an integral system of self-management in which co-ordinating self-management bodies would be created on the level of industries and also regionally. Membership in such associations would be obligatory (1965).

Mišić's system resembles the system of Higher Business Associations which existed in the two-year transitional period 1951–1952. A few years after self-management became operative, the present author suggested a somewhat different approach. A careful study of the economics of the oil industry showed that there was very little to be gained by competition and a lot to be achieved by a co-ordinated policy based on independent and competent research. I suggested that industries possesing similar characteristics establish common but independent economic-technological research institutes. The institutes would prepare alternatives for major policy decisions. The most acceptible alternative, perhaps modified in the process, would be chosen by the representatives of enterprises through some sort of self-management mechanism. The industrial research institutes would also serve as development planning institutions and as such would co-operate with territorial planning bureaus (Horvat, 1962c, ch. 24).

Self-management in enterprises is just one element in an integral system of social self-government. Pusić points out that such a system has three basic components: territorial (various levels of government); functional (enterprises and institutions, i.e., work organizations); and social (cultural, religious and other associations of individuals). Pusić is mostly concerned with the first component. He is thus the first among Yugoslav authors to study systematically the problem of the withering away of the state—generally considered utopian outside Yugoslavia. The state will wither away when government over individuals is replaced by the management of things. Engels took this famous phrase over from Saint-Simon. The latter, as well as other writers of his time, maintained that public administration was exclusively an instrument of power but that it was otherwise unimportant for the life of a nation. Marx and Engles argued with the first part of the statement, but regarded public administration as very important. Later an important duality appeared: public administration was no longer exclusively an instrument of power, but was also entrusted with various socially necessary activities: education, medical care, social welfare etc., basically differ from defense, police and judiciary. The monopoly of physical power might occasionally be useful is not at all necessary when social services are concerned. In socialism public administration without state political power becomes the question of the day. In other words, systematic planning and coordination of social services does not presuppose any longer the existence of a commanding center such as is political power (Pusić, 1968). The interest unions and the quasi-market, discussed in the section on institutional framework, represent an attempt to move in this direction.

Self-government is not a purely economic phenomenon. While economists are, naturally enough, primarily interested in economic aspects, other social scientists explore additional dimensions. Lj. Tadić, the political scientist, points out that Yugoslav self-government socialism is mostly confined to the economic sphere. It has been developed on the micro level without a corresponding reflection on the macro level, that of the global society (Simpozij, 1969, p. 55). S. Stojanović, the philosopher, maintains that without faster political democratization it is impossible to create self-government on higher levels of social organization (Simpozij, 1969, p. 34). R. Supek, the sociologist, explains that political pluralism does not mean a multi-party system which can also be bureaucraticized. In a self-government setting political pluralism means direct control of various centers of power. How this is to be achieved is an open problem. Supek expects a certain duality of power to develop at first, a combination of classical representative democracy and self-government.

Evidently, self-government is not a closed and complete system. Many questions are still open, many problems unresolved. The Yugoslav social laboratory is bound to be active for some time to come.

References

D. Anakiovski, "Foreign Trade in the Yugoslav Reform," *Yugoslav Survey,* 1969, *3,* 71–84.

D. Avramović, "Funkcija deviznog kursa u socijalističkoj privredi," *Ekonomist,* 1952, *3,* 3–31.

A. Bajt, "Osebni donodki in delovna storilnost," *Ekonomska revija,* 1956, vol. VII, 97–134.

———, *Raspodela nacionalnog dohotka i sistem ličnih dohodaka u našoj privredi,* Beograd 1962.

———, "Optimalna veličina investicija iz nacionalnog dohotka," *Ekonomist,* 1958, vol. XI, 79–91.

———, "Stopa rasta u nacrtu perspektivnog plana," *Ekonomist,* 1963, vol. XVI, 584–91.

———, "Izvori inflacije u razdoblju posle reforme," *Ekonomist,* 1967a, vol. XX, 141–46.

———, "Faktori dohotka i osnovne ekonomske zakonitosti u njegovoj raspodjeli u socijalističkoj tržišnoj privredi," *Ekonomist,* 1967b, vol. XX, 347–87.

———, "Yugoslav Economic Reforms, Monetary and Production Mechanism," *Economics of Planning,* 1967c, vol. VII, 201–18.

———, "Društvena svojina–koektivna i individualna," *Gledišta,* 1968, vol. XIX, 531–44.

———, "Fluctuations in Growth Rates in Post War Socialist Economies," *International Economic Seminar—CESES,* Balatonfüred 1969a.

———, "Privredna kretanja i ekonomska politika u 1969. i 1970. godini," in *Aktuelni problemi ekonomske politike Jugoslavije, 1969/1970,* Zagreb 1969b, pp. 5–17.

V. Bakarić, *Problemi zemljišne rente u prelaznoj etapi,* Zagreb 1950.

———, *Aktuelni problemi izgradnje našeg privrednog sistema,* Zagreb 1963.

———, *Aktuelni problemi sadašnje etape revolucije,* Zagreb 1967.

P. Basaraba, "Changes in the Organization and Management of Banks," *Yugoslav survey,* 1967, *4,* 77–81.

V. Begović, Dvije i po godine Petogodišnjeg plana," *Komunist,* 1949, *5,* 82–101.

E. Berković, "Differentiation of Personal Incomes," *Yugoslav Survey,* 1969, *1,* 81–90.

R. Bicánić, "Economic Growth Under Centralized and Decentralized Planning: Yugoslavia—A Case Study," *Economic Development and Cultural Change,* 1957, vol. V, 63–74.

———, *Ekonomska politika Jugoslavije,* Zagreb 1962a.

———, "The Threshold of Economic Growth," *Kyklos,* 1962b, vol. XV, 7–28.

———, "Centralističko, decentralističko ili policentričko planiranje," *Ekonomist,* 1963a, vol. XVI, 456–69.

———, "O monocentričnom i policentričnom planiranju," *Ekonomski pregled,* 1963b, vol. XIV, 469–528.

———, "Economics of Socialism in a Developed Country," *Foreign Affairs,* 1966, vol. 44, 633–50.

R. Bičanić, *Problems of Planning: East and West,* The Hague 1967.

D. Bilandžić, *Management of Yugoslav Economy: 1945–1966,* Beograd 1967.

———, "Odnosi izmedje samoupravljanja i rukovodjenja u poduzeću," in *Savremeno rukovodjenje i samoupravljanje,* Beograd 1969, pp. 67–96.

D. Bjelogrlić, "O nekim problemima društvenog usmeravanja privrede," in *Usmeravanje društvenog razvoja u socijalizmu,* Beograd 1965.

K. Bogoev, *Lokalne finansije,* Beograd 1964.

———, "Opšti prikaz fiskalnog sistema i fiskalne politika Jugoslavije," *Univerzitet danas,* 1966, *9–10,* 149–63.

———, "Stabilizaciona fiskalna politika," Ekonomist, 1967, vol. XX, 1–28.

D. Čalić, *Metodologija planiranja proizvodnje,* Beograd 1948.

D. Čehovin, *Ekonomski odnosi Jugoslavije s inostranstrom,* Beograd 1960.

F. Černe, *Planiranje in tržni mehanizem v ekonomski teoriji socijalizma,* Ljubljana 1960.

———, *Tržište i cijene,* Zagreb 1966.

———, "Poskus ekonomsko–logičnega testiranja sedem hipotez iz teorije dohodka," *Ekonomska revija,* 1967a, vol. XVIII, 12–29.

———, "O stabilizaciji in nihanjih v gospodarstvu," *Ekonomska revija,* 1967b, vol. XVIII, 212–29.

A. Cičin-Sain, *Devizni režim i konvertibilnost dinara,* Zagreb 1967.

———, "Problemi konvertibilnisti dinara," *Ekonomist,* 1968a, vol. XIX, 79–102.

———, "Fiksni ili fleksibilni kursovi," *Ekonomist,* 1968b, vol. XIX, 642–48.

M. Ćirović, *Novac i kredit,* Beograd 1966.

N. Čobeljić, *Politika i metodi privrednog razvoja Jugoslavije,* Beograd 1959a.

———, "Tri osnovna problema u teoriji razvoja nedovoljno razvijenih zemalja," *Ekonomist,* 1959b, vol. XII, 225–53.

N. Čobeljić, K. Mihajlović and S. Djurovic, "Problem našeg tržišta s naročitim osvrtom na tržište poljoprivrednih proizvoda," *Ekonomist,* 1954, *3–4,* 31–70.

——— and R. Stojanović, *Teorija investicionih ciklusa u socijalističkoj privredi,* Beograd 1966.

G. D. H. Cole, *Guild Socialism Re–Stated,* London 1920.

S. Dabčević *et al., O nekim problemima privrednog sistema,* Zagreb 1962. (Reprinted in *Ekonomski pregled,* 1963, *3–5.*)

S. Dabčević–Kučar, "Decentralized Socialist Planning: Yugoslavia," in E. E. Hagen, ed., *Planning Economic Development,* Homewood, Ill. 1963. pp. 183–222.

M. Dautović, "Economic Integration," *Yugoslav Survey,* 1968, *2,* 75–82.

J. Daviĉo, "Privredni problemi komune," *Ekonomist,* 1954, *3–4,* 185–95, discussion, 195–208.

D. Dimitrijević, "The Financial Structure in a Changing Economy: the Case of Yugoslavia," *Florida State University Slavic Papers,* 1968a, vol. II, 1–22.

———, "The Use of Flow–of–Funds Accounts in Monetary Planning in Yugoslavia," *Review of Income and Wealth,* 1968b, Series 14, 101–116.

J. Djordjević, "Teorijska i ustavna pitanja planiranja u Jugoslaviji," in *Usmeravanje* *društvenog razvoja u socijalizmu,* Beograd 1965. pp. 7–28.

———, "A Contribution to the Theory of Social Property," *Socialist Thought and Practice, 1966,* 24, 73–110.

M. Dobrinčić *et al., Privredni sistem FNRJ,* Zagreb 1951.

A. Domandžić, "Customs Tariff," *Yugoslav Survey,* 1966, 24, 3485–88.

E. Domar, "The Soviet Collective Farm," *American Economic Review,* 1966, vol. LVI, 734–57.

J. Dordević, *Sistem lokalne samouprave u Jugoslaviji,* Beograd 1957.

———, The Communal System in Yugoslavia," *Annals of Collective Economy,* 1959, vol. XXX, 169–207.

———, "A Contribution to the Theory of Social Property," *Socialist Thought and Practice,* 1966, *24,* 73–110.

A. Dragičević, *Potreban rad i višak rada,* Zagreb 1957.

———, **S. Štampar and B. Horvat,** *Naše Teme,* 1962, vol. VI, 872–94, 1318–33, 1487–1523; 1963, vol. VII, 99–100.

I. Drutter, "Uticaj koncentracije ponude na cijene i poslovni uspjeh privrednih organizacija," *Ekonomist,* 1964, vol. XVII, 697–700.

———, "Tržišni aspekti koncentracije," in Ekonomski institut, *Ekonomske studije 3,* Zagreb 1965.

———, "Sistem cijena i tržišnih odnosa," in ed. *Poduzeće u reformi,* Zagreb 1968. pp. 95–132.

D. Dubravčić, *Ponašanje samoupravnog Poduzeće u reformi,* Zagreb 1968. pp. 95–132.

———, "Prilog zasnivanju teorije jugoslavenskog poduzeća: Mogućnosti uopćavanja modela," *Ekonomska analiza,* 1968, vol. II, 120–27.

U. Dujšin, "Determinante izbora izmedu fiksnog i fleksibilnog kursa kod nas," *Ekonomist,* 1968, vol. XXI, 592–98.

J. Duričič, "Local Communities," *Yugoslav Survey,* 1965, vol. VI, 3287–300.

K. Džeba and M. Beslać, *Privredna reforma,* Zagreb 1965.

I. Fabinc, "Uloga carinske politike u zemljama u razvoju," *Medjunarodni problemi,* 1963, *4,* 27–39.

———, et al., "Problemi ekonomskih odnosa s inozemstvom," in ed. *Poduzeće u reformi,* Zagreb 1968a, pp. 133–216.

———, "Elementi programa zaštite jugoslavenske privrede," *Ekonomist,* 1968b, vol. XXI, 41–60.

A. Fiamengo, "Samoupravljanje i socijalizam," in Janičijević, ed., *Društveno samoupravljanje u Jugoslaviji,* Beograd, 1965, pp. 11–38.

W. Friedmann and L. Mates, eds., *Joint Business Ventures of Yugoslav Enterprises and Foreign Firms,* Belgrade 1968.

M. Frković, "Disparitet spoljnotrgovinskih kurseva u našoj privredi," *Ekonomist,* 1957, vol. X, 79–97.

A. Gams, "Društvena svojina i društveno usmeravanje," in *Usmeravanje društvenog razvoja u socijalizmu,* Beograd 1965, pp. 50–67.

M. Golijanin, "Credit and Money Control," *Yugoslav Survey,* 1967, *3,* 93–104.

D. Gorupić, and I. Perišin, "Proširena reprodukcija i njeno financiranje," *Ekonomski pregled,* 1965, pp. 109–30.

D. Gorupić, "Tendencije u razvoju radničkog samoupravljanja u Jugoslaviji," *Ekonomist,* 1967, vol. XX, 593–638.

———, "Samoupravno poduzeće i privredna reforma," in *Poduzeće u reformi,* Zagreb, 1968, pp. 3–26.

———, "Razvoj samoupravnih društvenih odnosa i samoupravno odlučivanje u privredi," *Ekonomiski pregled,* 1969, vol. XX, 1–26.

V. Guzina, "Medunarodni zajmovi i socijalistička izgradnja," *Komunist,* 1950, *6,* 21–79.

M. Hanžeković, *Problemi društvenih financija.* Mimeograph, Institute of Economics, Zagreb 1967a.

———, "Djelovanje porezne i monetarno-kreditne politike na stabilizaciju jugoslavenske privrede," *Ekonomist,* 1967b, vol. XX, 29–49.

V. Holjevac, *Kreditno–monetarni problemi.* Mimeograph, Institute of Economics, Zagreb 1967a.

——, *Kreditno monetarni problemi 1960–1967*. Mimeograph, Institute of Economics, Zagreb 1967b.

B. Horvat, "Još jedan prilog pitanju prelaznog perioda," *Ekonomist*, 1951, *5–6*, 45–56.

——, "O problemu rudničke rente," *Ekonomski pregled*, 1953, vol. IV, 253–57.

——, "The Optimum Rate of Investment," *Economic Journal*, 1958, vol. LXVIII, 747–67.

——, and V. Rašković, "Workers' Management in Yugoslavia: A Comment," *Journal of Political Economy*, 1959, vol. LXVII, 194–98. B. Ward, "Reply," 199–200.

——, "A Restatement of a Simple Planning Model with Some Examples from Yugoslav Economy," *Sankhya*, Series B, 1960, 29–48.

——, *Towards a Theory of Planned Economy*, Beograd 1964. (Serbo-Croatian ed., 1961.)

——, ed., *Uzroci i karakteristike privrednih kretanja u 1961. i 1962. godini*, Beograd 1962a.

——, "Raspodela prema radu medu kolektivima," *Našta stvarnost*, 1962b, vol. XVI, 52–66.

——, *Ekonomika jugoslavenske naftne privrede*, Beograd 1962c.

——, *Note on the Rate of Growth of the Yugoslav Economy*, Beograd 1963.

——, *Samoupravlenie, centralizm i planovanie*, Beograd 1964.

——, "The Optimum Rate of Investment Reconsidered," *Economic Journal*, 1965, vol. LXXV, 572–76.

——, "Prilog zasnivanju teorije jugoslavenskog poduzeća," *Ekonomska analiza*, 1967a, vol. I, 7–28.

——, "Jugoslavenski sistem samoupravljanja in uvoz tujega kapitala," *Ekonomska revija*, 1967b, vol. XVIII, 406–17.

——, *Ekonomska nauka i narodna privreda*, Zagreb 1968a.

——, "An Integrated System of Social Accounts for an Economy of the Yugoslav Type," *Review of Income and Wealth*, 1968b, Series 14, 19–36.

——, "Socijalistička robna proizvodnja," *Gledišta*, 1968c, vol. XIX, 1321–30.

——, *Ogled o jugoslavenskom durštvu*, Zagreb 1969a. Eng. ed., *An Essay on Yugoslav Society*, New York 1969a.

——, "Tehnički progres u Jugoslaviji," *Ekonomska analiza*, 1969b, vol. III, 29–57.

——, "Planning in Yugoslavia," paper presented at the conference on Crisis in Planning, University of Sussex, 1969c.

——, *Primjena medjusektorske analize u planskom bilanciranju privrede*, Beograd, 1969d.

——, et al., *Integrirani sistem društvenog računovodstva za jugoslovensku privredu*, Beograd 1969e.

——, *Privredni ciklusi u Jugoslaviji*, Beograd 1969. Eng. ed., *Business Cycles in Yugoslavia*, New York 1970

B. Ivanović, *Primena metoda I-odstupanja u problemima odredjivanja ekonomske razvijenosti*, Institut ekonomskih nauka, separat br. 13, Beograd 1964.

M. Janković, "Lični dohoci kao faktor podizanja životnog standarda," in *Obračun i raspodela osobnih dohodaka u radnim organizacijama*, Zagreb 1968, pp. 155–68.

B. Jelčić, "Poreski instrumenti kao instrument ekonomske politike," *Ekonomist*, 1967a, vol. 50–63.

——, *Problemi društvenih financija/prihoda/*. Mimeograph, Institute of Economics, Zagreb 1967b.

——, *"Ekonomski učinci oporezivanja prometa proizvoda*. Mimeograph, Institute of Economics, Zagreb 1967c.

——, "Poreska i budžetska politika," in *Aktuelni problemi ekonomske politike Jugoslavije 1969/1970*, Zagreb 1969.

B. Jelić, "Neki aspekti dejstva plana i tržišta u našoj privredi," *Ekonomist*, 1958, vol. XI, 183–201.

——, "Characteristics of the Yugoslav Economic Planning System," *Socialist Thought and Practice*, 1961, *1*, 59–81.

——, *Sistem planiranja u jugoslavenskoj privredi*, Beograd 1962.

B. Jovanović, "Reform of the Credit and Banking System," *Yugoslav Survey*, 1965, *22*, 3216–236.

P. Jurković, "Suština i značaj promjena u sistemu utvrdjivanja i raspodjele dohotka," *Ekonomski pregled*, 1969, vol. XX, 27–52.

E. Kradelj, *et al.*, *Razvoj privrede FNRJ*, Beograd 1956.

———, "Basic Principles of the New Constitution," *Yugoslav Survey,* 1962, *11,* 1529–56.

B. Kidrič, *Privredni problemi FNRJ,* Beograd, 1948.

———, "Kvalitet robno-novčanih odnosa u FNRJ," *Komunist,* 1949, *1,* 33–51.

———, *Privredni problemi FNRJ,* Beograd 1950a.

———, "Teze o ekonomici prelaznog u našoj zemlji," *Komunist,* 1950b, *6,* 1–20.

———, "O nekim teoretskim pitanjima privrednog sistema," *Komunist,* 1952, 41–67.

———, *Sabrana dela,* Knjiga III, Beograd 1960.

T. Konevski, *Fundamentalnost i razvojne smernice novog sistema finansiranja društveno-političkih zajednica,* Beograd 1968.

M. Korač, "Prilog pitanju o prelaznom periodu," *Ekonomist,* 1951, *3–4,* 37–46.

———, *Analiza ekonomskog položaja privrednih grupacija na bazi zakona vripednosti,* Zagreb 1968.

M. Košir, The *Kranj Commune,* Beograd 1966.

O. Kovač, *Uzroci i posljedice struckturne neravnoteže u platnom bilansu Jugoslavije.* Mimeograph, Institute of Economic Studies, Beograd 1966.

P. Kovač and Dj. Milijević, *Samoupravljanje proizvodjača u privredi,* Beograd 1958.

M. Kovačević, "Enterprise Rules and Regulations," *Yugoslav Survey,* 1969, *1,* 1–8.

S. Kraigher, "O politični ekonomiji v prehodnom razdoblju," *Ekonomska revija,* 1950, *1–2,* 9–46.

I. Lavrač, "Konkurencija i stimulacija u našem privrednom sistemu," *Ekonomist,* 1958, vol. XI, 601–19.

———, "Cena upotrebne vrednosti kapitala," *Ekonomska misao,* 1968, vol. I, 407–23.

B. Lazarević, "Turnover Tax," *Yugoslav Survey,* 1965, vol. VI, 3311–20.

G. Lêman, *Stellung und aufgaben der ökonomischen Unternehmungen,* Berlin 1967.

———, *Ungelöste Fragen in jugoslawischen System der Arbeiterselbstverwaltung,* Köln 1969.

F. Lipovec, "Razvoj profitne mere v sistemu samouprave delovnih kolektivov," *Ekonomska revija,* 1954, vol. V, 141–51.

S. Lovrenović, *Ekonomska politika Jugoslavije, Sarajevo* 1963.

G. Macesich, *Yugoslavia. The Theory and Practice of Development Planning,* Charlottesville, Va. 1964.

Lj. Madžar, "Jedna empirijska analiza stabilnosti spoljnotrgovinskih tokova," *Ekonomist,* 1968, vol. XXI, 580–87.

V. Majhsner, "Intervalutarni kurs i cene," *Ekonomski anali,* 1956, *3,* 186–204.

V. Majksner, "Intervalutarni kurs i cene,"

I. Maksimović, *Teorije socijalizma u gradjanskoj ekonomskoj nauci,* Beograd 1958.

———, "Razmišljanja o nekim teoretskim i idejnim pitanjima robne proizvodnje povodom našeg privrednog sistema," *Ekonomist,* 1964, vol. XVII, 209–26.

E. Mandel, "Yugoslav Economic Theory," *Monthly Review,* 1967, *11,* 40–49.

M. Marković, "Socijalizam i samoupravljanje," in *Smisao i perspektive socijalizma,* Zagreb 1965, pp. 54–71.

T. A. Marschak, "Centralized versus Decentralized Resource Allocation: The Yugoslav Laboratory," *Quarterly Journal of Economics,* 1968, vol. LXXXII, 561–87.

K. Marx, *Rani radovi,* Zagreb 1953.

M. Matejić, *Javne finansije,* Beograd 1958.

V. Medenica, "A Survey of the Major Results Achieved in the Implementation of Yugoslavia's 1966–1970 Social Plan," *Yugoslav Survey,* 1968, *4,* 27–46.

M. Mesarić "Prilog diskusiji o obliku gravitacione cijene u socijalistickoj, privredi," *Ekonomski pregled,* 1965, vol. XVI, 607–34.

———, *Planiranje privrednog razvoja,* Zagreb 1967.

———, "Uloga planiranja u jugoslavenskom privrednom modelu," *Ekonomist,* 1969, vol. XXII, 403–26.

P. Mihajlović and S. Tanović "Veza jugoslovenskog izvoza s konjukturom u svetu," *Ekonomist,* 1959, vol. XX, 45–79.

B. Mijović, *Novčana i kreditna politika,* Beograd 1967.

S. M. Milatović, *Poreski sistem,* Beograd 1967.

V. Milenković, "Spoljna trgovina," in *Razvoj privrede FNRJ,* Beograd 1956, pp. 399–399–419.

M. Miletić, "Da li je upravni odbor prevazidjen," *Direktor,* 1969, *9,* 56–60.

R. Milić, *Ekonomika FNRJ,* Beograd 1951.

N. Miljanić, *et al., Kreditni i finansijski sistem u Jugoslaviji,* Beograd 1956.

———, "Pilog izučavanju problematike novca," *Ekonomski pregled,* 1956, vol. VII, 12–24.

———, *Novac i kredit,* Zagreb 1964.

———, "Reguliranje monetarnog volumena u SFR Jugoslaviji," *Univerzitet danas,* 1966, 9–10.

D. Milivojević, *The Yugoslav Commune,* Beograd 1965.

Dj. Miljević, *Privredni sistem Jugoslavije,* Beograd 1965.

D. Miljković, "Komuna i društvena reprodukcija," in *Privredni sistem i ekonomska politika Jugoslavije,* Beograd 1961, p. 66.

D. M. Milojević, *Neoposredni porezi Srbije i Kraljevine Srba, Hrvata i Slovenaca,* Beograd 1925.

D. Mišić, "Sistem integralnog samoupravljanja u jugoslavenskoj privredi," *Ekonomist,* 1965, 289–312.

P. Mitić, "Ekonomske integracije, svjetsko tržište i Jugoslavija," *Gledišta,* 1969, vol. XX, 1073–86.

Z. Mrkušić, *Medunarodna trgovina i trgovinska politika,* Beograd 1963.

———, "Neka pitanja na alternativu: prilagodjavavje deviznog kursa—direktna kontrola," *Ekonomist,* 1967, vol. XX, 89–102.

———, "Problemi prilagojavanja deviznog kursa," *Ekonomska misao,* 1969, vol. II, 133–41.

———, *Spoljnoekonomska politika Trećeg svijeta,* in press.

E. Neuberger, "The Role of Central Banking under Various Economic Systems," in C. J. Friedrich and S. E. Harris, eds., *Public Policy,* Cambridge 1958, pp. 227–54.

———, "Centralization vs. Decentralization: The Case of Yugoslav Banking," *American Slavic and East European Review,* 1959a, vol. XVIII, 361–73.

———, "The Yugoslav Investment Auctions," *Quarterly Journal of Economics,* 1959b, 88–115.

D. Nikolić, ed., *Elementi metodologije planiranja dugoročnog privrednog razvoja,* Beograd 1964.

M. Novak, "O prelaznom periodu," *Ekonomski pregled,* 1952, vol. III, 203–13.

———, *Uvod u političku ekonomiju socijalizma,* Zagreb 1955.

———, *Organizacija poduzeća u socijalizmu,* Zagreb 1967.

S. Obradović, *Uvod u analizu spoljne trgovine,* Beograd 1962.

A. Papić, "Investment Financing in Yugoslavia," *Annals of Collective Economy,* 1959, vol. XXX, 208–31.

N. Pašić, *Javne korporacije u Velikoj Britaniji i drugim zapadnim zemljama,* Beograd 1957.

M. Pečujlić, *Klase i savremeno društvo,* Beograd 1967.

S. Pejovich, "Taxes and Pattern of Economic Growth: the Case of Yugoslavia," *Cahiers de l'ISEA,* 1964, G20/150 Suppl., 227–35.

———, *The Market-Planned Economy of Yugoslavia,* Minneapolis 1966.

J. Pelicon, "Sumarna ocena i neki problemi privredne suradnje SFRJ sa zemljama u razvoju u 1966-1967. godini," in *Privredni odnosi Jugoslavije sa zemljama u razvoju,* Ljubljana 1968, pp. 1–23.

M. Perović, "Još o prelaznom periodu," *Ekonomski pregled,* 1953, vol. IV, 29–42.

A. Perić, *Finansijska teorija i politika,* Beograd 1964.

I. Perišin, "Stabilizacija i monetarno-kreditna politika," *Ekonomist,* 1967, vol. XX, 103–120.

———, *Monetarno-kreditna politika,* Zagreb 1968.

———, "Antiinflatorna politika Jugoslavije poslije reforme," *Ekonomski pregled,* 1969, vol. XX, 497–530.

V. Pertot, *Yugoslav Foreign Trade,* Beograd 1960.

———, "Stabilizacija u uslovima disparitetnih odnosa troškova proizvodnje," *Ekonomist,* 1966, vol. XIX, 316–44.

M. Petrović, *Formiranje prihoda društveno-političkih zajednica u SR Srbiji i njihova raspodela izmedju republika, pokrajina i opština,* Beograd 1968.

J. Pokorn, "Razvoj našeg finansijskog sistema," *Finansije,* 1956, vol. XI, 1–10.

S. Popov, "Kretanje produktivnosti rada i ličnih dohodaka u pojedinim granama u periodu od 1952 do 1966. godine," in *Obračuni i raspodela osobnih dohodaka u radnim organizacijama,* Zagreb 1968, pp. 613–33.

Z. Popov, "Osvrt na kretanje privrednog razvoja u svetu," *Ekonomska analiza,* 1968, vol. II, 353–65.

M. Popović, "O ekonomskim odnosima izmedju socijalističkih država," *Komunist,* 1949, *4,* 89–146.

———, "O sistemu ekonomske i socijalističke demokratije u Jugoslaviji," *Komunist,* 1952, *3–4,* 1–14.

———, *Društveno ekonomski sistem,* Beograd 1964.

S. Popović, "Merenje dohotka i njegova raspodela," *Ekonomska misao,* 1968, vol. I, 424–36.

E. Pusić, *Samoupravljanje,* Zagreb 1968.

R. Radovanović, *Poreski sistem FNRJ,* Beograd 1953.

———, "Budžet u toku proteklih deset godina," in *Razvoj privrede FNRJ,* Beograd, 1956a, pp. 443–51.

———, *Oporezivanje dohotka privrednih poduzeća,* Beograd 1956b.

———, "Budgetary System and Budget Expenditure," *Yugoslav Survey,* 1962, vol. III, 1111–22.

M. Radulović, *Sistem i politika cijena u Jugoslaviji.* Unpublished doctoral dissertation, Titograd 1968.

V. Rajković, "Ocjena ostvarivanja privredne reforme i aktuelni problemi," in *Aktuelni problemi ekonomske politike Jugoslavije,* Zagreb, 1969/1970, pp. 21–48.

V. Rašković, "Osnovni idejni i politički problemi ličnog rada u sistemu društvenog samoupravljanja," in *Privatni rad: Za ili protiv,* Beograd, 1967a.

———, *Društveno samoupravljanje i raspodela prema radu u Jugoslaviji,* Beograd 1967b.

M. Samardžija, "Metodološke i društvene osnove teorije raspodele dohodka," *Gledišta,* 1968, vol. XIX, 124–49, 293–304.

J. A. Schumpeter, *Capitalism, Socialism, and Democracy,* New York 1950.

B. Šefer, "Tržište u posleratnom periodu," in *Razvoj privrede FNRJ,* Beograd 1956.

———, "Problemi i politika razvoja lične i društvene potrošnje," in J. Sirotković, ed., *Suvremeni problemi jugoslavenske privrede i ekonomska politika,* Zagreb 1965.

———, "Rasponi ličnih dohodaka, njihovo formiranje i tendencije," in *Obračun i raspodela osobnih dohodaka u radnim organizacijama,* Zagreb 1968a, pp. 421–38.

———, *Ekonomski razvoj Jugoslavije i privredna reforma,* Beograd 1968b.

M. Sekulić, *Primjena strukturnih modela u planiranju privrednog razvoja,* Zagreb 1968.

P. Sicherl, *et al., Izučavanje problema dopunskih sredstava republikama na trajnijoj osnovi.* Mimeograph, Institute of Economic Studies, Beograd 1968.

———, "Analiza nekih elemanata za ocenu stepena razvijenosti republika i pokrajina," *Ekonomska analiza,* 1969, vol. III, 5–28.

J. Sirotković, *Planiranje proširene reprodukcije u socijalizmu,* Zagreb 1951.

———, *Problemi privrednog planiranja u Jugoslaviji,* Zagreb 1961.

———, *Planiranje u sistemu samoupravljanja,* Zagreb 1966.

B. Šoškić, "Rast proizvodnje i zaposlenosti i mere ekonomske politike," *Ekonomist,* 1969a, vol. XXII, 143–55.

———, "Povećanje zaposlenosti u našem sistemu tržišne privrede," *Ekonomska misao,* 1969b, vol. II, 79–92.

S. Srdar, *Da li FNR Jugoslavija postaje agrarnouvozna i industrijski izvozna zemlja,* Zagreb 1953.

V. Stanovčić and A. Stojanović, eds., Books I and II, *Birokratija i tehnokratija,* Beograd 1966.

R. M. Stevanović, *Novčani i kreditni sistem,* Beograd 1954.

S. Stojanović, "Etatistički mit socijalizma," *praxis,* 1967, vol. III, 30–38.

R. Stojanović, "Stopa rasta socijalističke privrede," in R. Stojanović, ed., *Savremeni problemi privrednog razvoja u socijalizmu,* Beograd 1960.

M. Sukijasovič and Dj. Vujačić, *Industrial Cooperation and Joint Investment Ventures Between Yugoslav and Foreign Firms,* Beograd 1968.

P. Sweezy, "The Transition from Socialism to Capitalism?", *Monthly Review,* 1964, vol. 16, 569–90.

Z. Tanić, ed., *Radničko samoupravljanje; razvoj i problemi,* Beograd 1963.

T. Tišma, *Javne finansije,* Zagreb 1964.

M. Todorović, *Oslobodjenje rada,* Beograd 1965.

T. Tomić, "Dosadašnji razvoj raspodele ličnih dohodaka u SFRJ," in *Obračun i raspodela osobnih dohodaka u radnim organizacijama,* Zagreb 1968, pp. 3–22.

M. Toroman, "Oblici društvene svojine." Paper presented at the Symposium on Social Ownership, Serbian Academy of Science and Art, Beograd, Sept. 20–22, 1965.

M. Trklja, *Kamata na fondove u privredi.* Mimeograph, Institute of Economic, Zagreb 1968.

S. Turčinović, "Financing Socio-Political Units," *Yugoslav Survey,* 1968, *2,* 59–74.

R. Uvalić, "Zakon vrednosti i njegovo korišćenje u planiranju narodne privrede," *Ekonomist,* 1948, *1,* 20–27.

———, "O nekim principima našeg privrednog sistema i problemi njihove primene," *Ekonomist,* 1954, *3,* 5–17.

———, "Funkcije tržišta i plana u socijalističkoj privredi," *Ekonomist,* 1962, vol. XV, 205–19.

———, "Trojna ekonomska suradnja Jugoslavije, Indije i UAR–a," in *Privredni odnosi Jugoslavije sa zemljama u razvoju,* Ljubljana 1968, pp. 128–46.

F. Vasić, "Investment in the Post–War Period," *Yugoslav Survey,* 1963, *15,* 2153–172.

Z. Vidaković, *Promene u strukturi jugoslavenskog društva i Savez Komunista,* Beograd 1967.

D. Vojnić, "Investiciona politika i sistem proširene reprodukcije," in *Aktuelni problemi ekonomske politike Jugoslavije 1969/1970,* Zagreb 1969, pp. 75–92.

M. Vučković, *Naš novi planski i finansijski sistem,* Beograd 1952.

———, "Preduzeće i kredit," *Ekonomski anali,* 1956, vol. II, 166–85.

———, *Kreditni sistem u FNRJ,* Beograd 1957.

———, "The Recent Development of the Money and Banking System of Yugoslavia," *Journal of Political Economy,* 1963, vol. LXXI, 363–77.

———, "Dosadašnja inflaciona kretanja u Jugoslaviji," *Ekonomist,* 1967, vol. XX, 121–140.

D. Vuković, "Price Formation and Social Price Control," *Yugoslav Survey,* 1968, *1,* 51–58.

R. Vuksanović, "Credit and Money," *Yugoslav Survey,* 1966, vol. VII, 3461–74.

H. M. Wachtel, *Workers' Management and Wage Differentials in Yugoslavia.* Unpublished doctoral dissertation, Univ. Mich. 1969.

B. Ward, "Workers' Management in Yugoslavia," *Journal of Political Economy,* 1957, vol. LXV, 373–86.

———, "The Firm in Illyria: Market Syndicalism," *American Economic Review,* 1958, vol. XLVIII, 566–89.

———, "Marxism-Horvatism: A Yugoslav Theory of Socialism," *American Economic Review,* 1967, vol. LVII, 509–23.

M. Žiberna, "Neki problemi ekonomskih odnosa s evropskom ekonomskom zajednicom," *Medunarodni problemi,* 1969, vol. XXI, 51–66.

J. Županov, "Radni kolektiv i ekonomska jedinica u svjetlu organizacione teorije," *Ekonomski pregled,* 1962, vol. XIII, 143–69.

———, *O problemima upravljanja i rukovodjenja u radnoj organizaciji,* Zagreb 1967a.

———, "Proizvodjač i riziko—Neki socijalno-psihološki aspekti kolektivnog poduzetništva," *Ekonomist,* 1967b, vol. XX, 389–408.

Ekonomist, "Diskusija ekonomista o prednacrtu ustava," 1962, vol. XV, 439–517.

Finansijski institut, *Finansijski sistem FNR Jugoslavije,* Beograd 1949.

Informativni jriručnik o Jugoslaviji, "Izvještaji savezne planske komisije," October 1948.

Institut društvenih nauka, *Koncepcija i verifikacija specifične cene priozvodnje u jugoslavenskoj privredi 1964. i 1965,* Beograd 1968.

Institut ekonomskih nauka, *Nauka i ekonomska politika,* Beograd 1968a.

———, *Sumarna analiza privrednih kretanja i prijedlozi za ekonomsku politiku,* Beograd 1968b.

———, *Ocjena ekonomske sitiacije i predvidjanja daljeg razvoja,* Beograd 1969.

Institut za spoljnu trgovinu, *Analiza devizne reforme iz 1961,* Beograd 1964.

International Labour Office, *Workers' Management in Yugoslavia,* Geneva 1962.

Jugoslavenski institut za ekonomska istraživanja, *Sumarna analiza privrednih kretanja i prijedlozi za ekonomsku politiku,* Beograd 1968.

Narodna banka Jugoslavije. *Novčano-kreditna politika i stabilnost dinara.* Mimeographed paper, Beograd 1965.

Program Saveza Komunista Jugoslavije, Beograd 1958.

Savezna skupština, *Osnovni problemi daljneg razvoja privrednog sistema,* Beograd 1964.

———, *Privredne reforme,* Beograd 1965.

———, *Osnove sistema društvenog planiranja,* Beograd 1966a.

———, *Devizni i spoljnotrgovinski režim,* Beograd 1966b.

Savezni zavod za statistiku, *Jugoslavija 1945–1964,* Beograd 1965.

Savjetovanje jugoslavenskih ekonomista, Zagreb 17–19 januara 1963, "Aktuelni problemi privrednog razvoja i privrednog sistema Jugoslavije," *Ekonomist,* 1963, *1.*

Savjetovanje jugoslovenskih ekonomista, "Problemi teorije i politika cena," *Ekonomist,* 1964, vol. XVII, 499–792.

———, Ljubljana, 9–11 marta 1967, "O uslovima stabilizacije jugoslavenske privrede," *Ekonomist,* 1966, *1–4,* 1967, *1–2.*

Simpozij jugoslavenskih-čekoslovačkih filosofa, "Savremeni trenutak socijalizma," *Filosofija,* 1969, vol. XIII, *2,* 1–98.

Yugoslav Survey, "Resolution of Federal Assembly on the Guidelines for Drawing up Yugoslavia's Social Plan for the 1964–1970 Period," 1964, 2703–16.

U. S. Congress, Senate Subcommittee on Antitrust and Monopoly of the Committee on the Judiciary, *Problems of Market Power and Public Policy in Yugoslavia,* by J. Dirlam, 90th Cong., 2nd sess., 1968, pp. 3758–85.

The American Economic Review

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• THE AMERICAN ECONOMIC REVIEW, ncluding four quarterly numbers, the Proceedings of the annual meetings, and *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September and December.

• Membership dues of the Association are $20.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature*. Subscriptions by nonmembers are $20.00 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $4.00 each. Each order for copies of either publication must also include a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

• Correspondence relating to the *Paper and Proceedings*, the *Directory*, advertising, permission to quote, business matters, subscriptions, membership and changes of address may be sent to the secretary, Rendigs Fels, 1313 21st Avenue, South, Nashville, Tennessee 37212. To be effective, notice of change of address must reach the secretary by the 1st of the month previous to the month of publication.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

Officers

President

JAMES TOBIN
Yale University

President-Elect

JOHN KENNETH GALBRAITH
Harvard University

Vice-Presidents

JAMES M. BUCHANAN
Virginia Polytechnic Institute
FRANCO MODIGLIANI
Massachusetts Institute of Technology

Secretary-Treasurer and Editor of Proceedings

RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

MARY JEAN BOWMAN
University of Chicago
CHARLES L. SCHULTZE
Brookings Institution
ROBERT DORFMAN
Harvard University
ARNOLD C. HARBERGER
University of Chicago
ROBERT EISNER
Northwestern University
JOHN R. MEYER
Yale University

Ex Officio Members

WILLIAM FELLNER
Yale University
WASSILY LEONTIEF
Harvard University

The 1971 Report of the President's Council of Economic Advisers: Inflation and Recession

By MARTIN J. BAILEY*

This year's Report presents a picture of turbulence and cautious anxiety seldom seen in these documents. Naturally, any recession is awkward for a national Administration and for its Council of Economic Advisers. To find one quite this awkward we must go back to 1949–50, when the then semiannual reports pointed exactly the wrong way three times in succession. (Shortly thereafter, the practice of having semiannual economic reports ended.) Moreover, we would have to try even harder to find an Economic Report as candid as this one about its problems. It presents alternative data for almost every key statistic, so that critics may freely draw their own conclusions from the Report's data if they wish.

I

The main source of unusual concern is inflation persisting at almost 6 percent during rising unemployment; both the inflation and the unemployment exceed the Council's previous forecasts and best hopes. Virtually all of this year's interesting policy issues and interpretive questions flow directly from these surprises. Even without them, any Administration has a painful choice among the points on each year's Phillips curve. Now that the curve turns out to offer even worse combinations than the current Administration had hoped, its choice is so much the more painful. Egged on by critics, the Administration is understandably tempted by the hope that it can get a better deal; by complaining, threatening, or intervening in private wage and price policies, to reduce the measured rate of inflation without adding to unemployment.

Behind these policy problems lie a series of interpretive issues, some central and some of academic flavor. The most central such issue is the location of the Phillips curve for 1971, and how far we dare move on it without creating an unacceptable curve for 1972. Associated with that, the Report asks rhetorically (p. 60), why has the choice proved so painful—"why is the inflation so stubborn?" In addition, it considers a number of detailed interpretive questions such as how to measure the impact of lowered defense spending, how to measure the effect of the auto strike in the fourth quarter of 1970, how to measure the rate of inflation, and how to measure the impact of fiscal policy. On this last question, it confers establishment recognition on the full employment budget.

II

The new official status of the full employment budget caps the change of official outlook toward fiscal policy, over the past forty years, from a strict "balance the budget" doctrine to the view that fiscal policy is a proper and useful instrument of economic stabilization.[1] Throughout official Washington, including even the Treasury, it is now not only proper to welcome a budget deficit as a desirable moderating influence during a recession, but it is also accepted as proper to enlarge the deficit as a positive countermeasure. These viewpoints have received official support in varying degrees for decades, but are now firmly established as doctrine. However, the commitment to fiscal activism is bounded.

---

* Professor of economics, University of Rochester.

[1] See Herbert Stein. Also see the discussion on pp. 171–75 of my *National Income and the Price Level*.

> The President's Budget for 1972 is based on the principle that expenditures should not exceed the revenues that the tax system would yield under conditions of full employment. This is an important principle. It permits the federal budget to support the economy when the economy is weak, by allowing the federal budget to move into a deficit under those conditions. But it retains the fiscal discipline of budget balancing by drawing a line beyond which expenditures may not go without tax increases. [p. 85]

The budget chosen to carry out this principle adds $17 billion to expenditures on a national income accounts basis between calendar 1970 and calendar 1971. On the other side, in addition to concurring in letting the income tax surcharge expire, the Administration liberalized depreciation allowances for a combined effect of over $11 billion (though less than $5 billion of that is a revenue reduction in 1971 compared to 1970). The restrained nature of fiscal activism comes through even more clearly when we note that the extra expenditures are more than accounted for by transfer expenditures, such as social security and revenue sharing. Federal purchases of goods and services are supposed to decline by $1.9 billion on a calendar-year basis, being dominated by the decline in defense spending. Thus the current Administration relies mainly on the Friedman-CED idea of letting fiscal policy remain automatic over the business cycle. The proposed discretionary adjustments are small, especially as compared to what could have been expected had neither recession nor inflation been a problem. Its defense of this caution is that big fiscal moves are hard to reverse.

III

How serious or surprising we view the current recession and inflation depends how far we look back and especially how we interpret past experience. In some respects the inflations of 1946–49 and of fall 1950 are the most comparable of the past fifty years, both in severity and in their association with war. In both cases inflation suddenly ran out of steam on its own, or at least presented nothing like the present policy problem. Another comparison, used in the Report, is that with 1955–57 (see p. 76). The inflation of that period continued into the recession of 1958, dropping by only about one percentage point on a full-year basis in the first year of recession compared to the year before, using any of the usual indexes. Then, in the following year of recovery, inflation stopped and so did the prevalent talk of economic power, wage-price spirals, inflation as a permanent way of life, etc. It is natural to expect that in 1971 there will be at least a partial turnabout of the same kind. However, the Report's confidence on this score goes with thinly masked anxiety, because there is no way to read the price statistics to find anything like a one percentage-point reduction in the rate of inflation in 1970 compared to 1969, on a full-year basis.

Quarterly price data and the auto strike give the Report a few straws to grasp at, but even with the farm sector left in, the annual price data present a gloomy picture. What should we infer? Within the relevant range of experience, we have two past war-associated inflations that ended suddenly, one peacetime inflationary episode that subsided slightly in the downturn and then stopped in the recovery, and the current case. These offer few data points for statistical inference or prediction. Accordingly, the Report combines its optimism with all due caution, saying that "There is a feasible path . . . that would see the unemployment rate reduced to the $4\frac{1}{2}$ percent zone by the second quarter of 1972 and the inflation rate, as measured by the *GNP* deflator, declining to approach the 3 percent range at the same time" (p. 78).

The policy instruments that are to keep the economy on this feasible path are the fiscal policy described above, a 6 percent rate of growth in the stock of currency plus demand deposits, inflation alerts, increases in oil import quotas, and intervention in the construction industry.

The key to this posture is a clear belief that the Phillips curve will shift. Past research on the inflation-unemployment choice suggests that the curve has two fixed branches, where the higher, less attractive one applies in an upswing and the lower, more attractive one applies in a downswing. The authors of the Report hold a different view:

> Confining the economic expansion to a pace which would keep unemployment about where it now is, in the neighborhood of 5.5 to 6.0 percent, would permit a significant decline in the rate of inflation during 1971 and 1972 . . . on the other hand, trying to restore what has been commonly regarded as full employment—a 4 percent unemployment rate—within the present planning period that extends to the end of fiscal year 1972 would entail risks on the inflation side. Although this latter path might be consistent with some further reduction of the inflation rate, there is a serious risk that the inflation rate would start rising again if the 4 percent unemployment rate were approached as rapidly as such timing would imply. [pp. 77–78]

We might call this a decelerationist view of the Phillips curve, the logical complement to the accelerationist view—the same idea applied to recession. This idea has less than universal acceptance, so it faces a test. Just as the past two years were supposed to have presented a crucial test of the monetarist versus the contrary view of aggregate real output (although in fact, with the war winding down, it wasn't too clean a test), the next year or so presents such a test of the orthodox Phillips curve view versus the accelerationist-decelerationist view.

We might claim, indeed, that this test has already occurred. Five years ago no one drew a Phillips curve through the almost 6 percent inflation at 6 percent unemployment that we have lately experienced. The rising inflation of the late 1960's while the unemployment rate fluctuated between 3.7 and 4.0 percent, points to a similar thought. However, there is the possibility of "structural change" or permanent upward shift of the curve, often suggested (as it was in the late 1950's), and also alluded to in the Report (pp. 60–66). If the change is permanent, there will be no downward shift in the curve in 1971 and 1972; inflation could in that case be slowed only by still higher unemployment. The test will be reinforced by the distinction between the downswing branch and the upswing branch of the curve. But no movement along any standard Phillips curve permits the rate of inflation to fall *concurrently* with a fall in percentage unemployment. Along the standard path, inflation subsides when and only when unemployment is rising; then when the economy turns around, they both reverse along the less favorable path. The Report flatly predicts a different course, saying that inflation will subside *after* recovery begins and unemployment starts declining. Time will tell.

IV

The policy measures noted above—a slightly activist fiscal policy, a monetary policy that, if carried out as advertised, will be more restrictive than it has been in past recoveries, and minor interventions that so far are hard to distinguish from those used in past inflations, recessions, and recoveries, seem anything but exciting in the face of an unusual problem. Are they enough?

By way of perspective, this recession opposes the sharpest sustained inflation since the late 1940's. In the past, the official view of the Federal Reserve Board has been that the sharper and more sustained the inflation it might permit with loose monetary policy, the more severe would be the recession that inevitably followed. On the face of it, that sounds reasonable. Yet in fact the economy seems to be climbing down from this sharp inflation with a recession that so far bears comparison with the mildest postwar recession, that of 1954. As a deliberately engineered attempt to cool off the economy, this recession may disappoint and embarrass the Administration, but do we have any technique that we can rely on to do better? Whenever fiscal and monetary restraint combine to choke back aggregate demand, as they did in 1969–70, business sentiment resists the restraint at first, and then turns around sharply in a virtual stampede. If there is any way to avoid this stampede, no one has yet demonstrated it. We may argue on general principles that bringing monetary expansion to a full halt, as the Federal Reserve System did in the latter half of 1969, overreacts to previous expansion and provokes too sharp a business contraction. Perhaps holding the expansion rate of the money stock to a rate of 4 or 5 percent per year would have provided enough stringency, along with the reduction in defense spending, to end the inflation; if so, perhaps any attendant recession would have been even milder than this one. And if so, the missed opportunity lies at the door of the Federal Reserve System.

In fairness, however, it is hard to be confident that anything less than the current mild recession could check the rampant and growing inflationary sentiment of recent years. A Mundell model for dynamic stabilization tells us that it is technically possible to end an inflation smoothly, using monetary swings whose immediate effects offset all lagged reactions within the system (see Robert Mundell, p. 97). Assuming that the system's parameters are reliable and known, and that it has the main features of Mundell's model, the authorities could choose any path they like for the price level and for real output; in particular, they could stop inflation cold and hold everything level. A critical assumption for this result is that the link between the quantity of money and the price level in the current period, expressed through demand and supply curves for money but actually operating through commodity and factor markets, is strong enough to overcome all lagged effects (including its own) without generating an explosive policy path. Even more critical is the assumption that the system's parameters are reliable and known. When the lags are unpredictable, depending in part on unique circumstances of the current period, the authorities must necessarily settle for something less than perfect stabilization. Moreover, we have no sure way to judge whether one policy path would work out better than some other in any particular instance such as the current inflation.

Restricting the rate of monetary growth below that of 1966–68 would necessarily provoke a reaction in the real economy eventually. If the restriction were milder than a dead stop, we can be confident that the reaction would have come later; but can we be sure that it would have been less? Comparisons with particular past recessions may give hints but can't give us clear cut answers, because of the distinctive features of each and the small number of cases. Thus, after noting that this recession is at the mild end of the range of experience of recent decades, any further judgement must be speculative.

Speculation does not end there; similar remarks apply to the other controversial areas of stabilization policy. The idea of contracyclical expenditure swings still commands wide support, although its supporters seem more alert than they used to be to the risk of too much, too late. By the time any ordinary recession goes far enough to leave no reasonable doubt of what it is, recovery is so near that the fiscal remedy, after normal delays of appropriation and execution, is likely to prove ill-advised. This point seems especially relevant now, when the current recovery, however shaky, seems definitely started. But we still have both inflation and unemployment, and the prospect for 1972 is that we will then still have more of both than we would like. The Report says we must accept that, although its expressed hope of a $4\frac{1}{2}$ percent rate of unemployment by the second quarter of 1972 implies faster recovery than most outside observers can believe.

The question is not whether our only choice is to accept prospective rates of inflation and unemployment, but whether it is wise to do anything *but* accept them. The Congress could legislate a stability of sorts; it could pass a law with the two provisions (a) that prices and incomes shall stop rising, and (b) that a job shall be available for every person seeking work. If we and they were willing to take all the consequences, the authorities could even execute this law. Without belaboring the point, I assume that we are unwilling to take all the consequences, and so reject the full extremity of such a law. But then, what about going part way? Could a small move, in a direction that would be unprofitable if pushed too far, yield gains that offset the costs? Or would it yield only costs and no gains? These are divisive questions, the evidence on which is so sparse and tenuous that everyone can, and does, pick his own Bayesian druthers to answer them. Those who favor intervention cite the price controls of 1951, which incidentally was also the year of the accord between the Federal Reserve and the Treasury; they remain discreetly silent about 1946–47, when controls failed to stop an inflation that later ended in a sharp recession. But what is very hard to understand is a position that says that a full battery of controls won't work, and then proposes partial controls. The Report comes perilously close to this latter position, one hopes as a compromise that will hold off a stampede into something worse (pp. 79–82).

V

$1,065? The Report does the best job it can of hewing the line gracefully, but with due professional caution.

> There is a considerable body of opinion that expects the gross national product for 1971 to be in the range between $1,045 billion and $1,050 billion, which would be an increase of 7 to $7\frac{1}{2}$ percent above that for 1970. This is a possible outcome. However, it would seem more likely that with present policies the outcome would be higher than that and could be as high as $1,065 billion . . . Nevertheless, although emerging information may later suggest a different view, the figure of $1,065 billion for the *GNP* in 1971 is an appropriate intermediate target. . . . [pp. 84–85]

With caveats delicately repeated, the Report's inspirational message says the magic number six times in the span of four paragraphs. Inasmuch as the Report properly mentions the wide range of uncertainty, we can only underline the possibility that its span may be even wider than the cited range of opinion. Although the signs since the beginning of this year have indicated revival, it is still too soon to be sure, especially as the signs are no stronger now than they were then. It is almost too late now for the sharp recovery the Report calls for. Nevertheless, we wish them luck.

As one who has leafed through a number of past such volumes without enthusiasm, I found this one refreshing in its clarity and professional balance. Perhaps next year's will make even better reading.

REFERENCES

M. J. Bailey, *National Income and the Price Level*, 2d ed., New York 1971.

R. A. Mundell, "Growth, Stability, and Inflationary Finance," *J. Polit. Econ.*, Apr. 1965 *73*, 97–109.

H. Stein, *The Fiscal Revolution in America*, Chicago 1969.

U.S. Council of Economic Advisors, *Economic Report of the President*, Washington 1971.

The 1971 Report of the President's Council of Economic Advisers: Inflation and Recession

By ROBERT EISNER*

Unemployment is too high. Output is too low. Price inflation is too great. Our economic policies have failed because: 1) the President (as well as the Congress), wedded to certain political goals, have failed to follow our advice; 2) we have not always known what advice to give. You will not find precisely those words in the 1971 Report of the Council of Economic Advisers, or anything like it in previous reports, and more is the pity.

This year's Annual Report should actually receive reasonably high marks from the apostles of the conventional wisdom. It is a balanced document, economically and politically. At first glance, there are twin objectives, or rather the compositive objective, ". . . to root out inflation without consigning our free economy to the stagnation of unemployment" (p. 3). And obeisance is done to the twin tools of fiscal and monetary policy, ". . . partly because it was not certain that primary reliance on either alone could be counted on to yield the desired overall results" (p. 23).

There is hardly ever any solid hint that the goals of the Council differ from those of the Administration. The aim is

> . . . a path that would see the unemployment rate reduced to the $4\frac{1}{2}$ percent zone by the second quarter of 1972 and the inflation rate, as measured by the *GNP* deflator, declining to approach the 3 percent range at the same time.
>
> . . . .
>
> The general goal . . . is that the rate of unemployment should decline as fast as is consistent with a reasonably steady and durable decline of the rate of inflation. [p. 78]

The reader thus senses, here and elsewhere, that despite the 6 percent unemployment rate at the end of 1970 (and again in March 1971), checking the rate of inflation has prime place in the Council's thinking. We are told that "concerted policies of expansion are needed now to lift the economy fast enough to make rapid progress toward full employment, and these needs will be fully met" (p. 7). But we are also told that "inflation . . . has prevented an all out attack on unemployment and . . . contains the possibility of frustrating the recovery policies which are being adopted" (p. 28).

Unemployment problems are repeatedly related to reduction of employment attributable to Department of Defense expenditures, a decline of ". . . nearly 1.8 million workers from its highest recent level in fiscal year 1968 to fiscal 1971" being cited (p. 43). How can a liberal critic argue with that? Indeed, the source of our troubles is to be found elsewhere than in current policies; Administration measures ". . . will not relieve the country of the consequences of past errors which have caused us to live for a longer time with both more unemployment and more inflation than anyone would like. They will leave us dependent upon a course of steady but not excessive economic expansion . . ." (p. 82). That is apparently a Council plea of innocence to go with the acknowledgement that "... 1970 ended with unemployment in the neighborhood of 6 percent and output in the fourth quarter about 6½ percent below its potential" (p. 82). The Council found a further culprit in the General Motors strike, which not only served to lower sharply fourth quarter *GNP*, but because of the interaction of weighting factors and price movements, brought about a spurious spurt in the *GNP* price deflator.

---

* Professor of economics, Northwestern University, and research associate, National Bureau of Economic Research.

The delicate—or indelicate—balance of economics and politics is interestingly displayed in the Report's handling of the notorious $1,065 billion projection of 1971 *GNP* along with the defense of a full employment balanced budget. The Report acknowledges,

> There is a considerable body of opinion that expects the gross national product for 1971 to be in the range between $1,045 billion and $1,050 billion.... This is a possible outcome. However, it seems more likely that with present policies the outcome would be higher than that and could be as high as $1,065 billion.
>
> ....
>
> It is reasonable to expect that with an increase of the *GNP* to $1,065 billion in 1971, the rate of price increase would be declining through the year, the unemployment rate would also end the year significantly lower than at the end of 1970, and real output would show a strong gain.
>
> ....
>
> If the rise in the money stock were to continue at the 1970 rate, the ratio of money to the *GNP* would then decline at about the average rate of the period 1952–70. Although this is a possible development, it is not a certainty. [p. 84]

And finally, a $1,065 billion *GNP* "... is feasible, and its realization with the proposed budget and complementary monetary policy is a reasonable expectation.... Still further increases of expenditures beyond this Budget or cuts in taxes would not have been consistent with fiscal discipline" (p. 85).

So what is the meaning of "a possible development" which "is not a certainty"? Does the Council accept the crude quantity theory of money approach which underlay the $1,065 billion projection? Apparently not really! For budget or fiscal policies are important. But what is this "fiscal discipline"? If it becomes clear to all that the "reasonable expectation" of a *GNP* of $1,065 billion is not being realized, although according to the Council such a rate of *GNP* would be consistent with desirably declining rates of price increase and unemployment, does the Council insist nevertheless that "fiscal discipline" would still make further increases of expenditures or cuts in taxes undesirable?

This leads us back to the Council's curious defense of a balanced full employment budget. In the President's Report (in the same volume) is found the apparently basic principle, "... except in emergencies, ***expenditures should not exceed the revenues that the tax system would yield when the economy is operating at full employment***" (p. 6, italics in original). But then in the Council Report we find,

> The absolute level of the full employment surplus or deficit is of limited significance for indicating how much restraint or stimulus the budget would exert on the economy if it followed the full employment path, or indeed for indicating which of these directions its influence would take. Changes in the full employment surplus from period to period are much more important indicators of how much fiscal policy is moving toward contraction or expansion. [p. 72]

Let us take the Council at its word. What is the change in the full employment surplus from fiscal 1971 to fiscal 1972? In fact, a chart presented with the President's Budget highlights a $1.4 billion surplus in the unified budget on a full employment basis in fiscal 1971 as against a $0.1 billion surplus originally estimated for the 1972 full employment budget. But it is clear that the 1971 Budget was not sufficiently expansionary to prevent

—or it was sufficiently constraining to achieve—the 6 percent unemployment and the first annual decline of real *GNP* in a long time. How then can the proposed full employment budget which differs by being only $1.3 billion less in surplus than the 1971 Budget prove sufficiently stimulatory or, in the words of the President's Budget message, "self-fulfilling"? The Council is perhaps to be commended for bringing the President around from his earlier expressed views that all Federal expenditures in excess of revenues should, as a first approximation, be viewed as sinful. But what is the new fiscal discipline that accepts the constraint of a balanced full employment budget regardless of the consequences for employment and output? (It may be noted that the calendar 1970 national income accounts full employment budget was actually $6.7 billion in surplus!)

President Nixon has been reported to have become a "Keynesian." I would wish that he were a better one. His position now, and that apparently of the Council, remains more "classical" than one of 1948 that "the budget principle might be *either* that the hypothetical [tax] yield . . . at a level of income corresponding to reasonably full employment at a predetermined price level . . . should balance government expenditure . . . or that it should lead to a *deficit sufficient to provide some specified secular increase in the quantity* of money" (italics added). This view was then expressed by Milton Friedman (1953, p. 137).

The President's Report, as noted above, called for "concerted policies of expansion," but the major new item of stimulatory fiscal policy cited is a $2.7 billion reduction in taxes resulting from arbitrarily "liberalized" depreciation allowances. We have here a large topic in itself and one of considerable controversy and litigation as I write (May 1971), but there can hardly be any student of business investment who can argue that this measure would have any prompt effect in stimulating equipment expenditures and many would argue, as I do, that it is a particularly poor way of stimulating investment even in the long run, with relatively little "bang for the buck." For capital expenditures are undertaken by business in anticipation of the profits to be realized from them. Giving business such a major tax concession will have little effect on the profitability of new or additional investment, particularly in a period of substantial idle capacity. It is hard to reconcile a $2.7 billion initial sacrifice of tax revenues—and growing amounts in the future—from our most capital-intensive companies, and partly as a consequence our generally largest companies, with the crying public and private needs that we are unable to meet as we restrict expenditures to full employment revenues. One cannot avoid the suspicion that the question of fiscal discipline, as well as the entire set of Administration policies in which the Council apparently acquiesces, seems to reflect certain ideological preferences as much as general economic principles. The economy is to be stimulated by a cut in *business* taxes and that, as noted, most beneficial to large, capital-intensive firms. Unemployment, which bears most heavily on the poor, the young, and the black, is to be combatted. But clearly this will be subject to the constraint of slowing the rate of price increase, usually a more dominant concern of the middle class and the elderly.

It may well be that the rate of inflation is finally slowing. Yet suggestions to this effect in the President's Report are uneasily reminiscent of the "prosperity-is-just-around-the-corner" statements of over a generation ago. The Council Report declares, "The rate of inflation for the most comprehensive price measure, the *GNP* deflator, reached a peak in the first quarter of 1970; in the next three quarters it recorded smaller rates of increase than in the first" (p. 49). But the successive rates of increase were 6.4 percent, reflecting a Federal government pay raise in that first quarter, 4.3 percent, 4.6 percent, and 5.7 percent (the last figure subsequently revised to 5.9 percent). It is hard to find a checking of the rate of inflation in these successive quarterly changes of the deflator, but the Council makes a valiant effort, calculating a price deflator excluding autos, which shows successive increases for the four quarters of 1970, of 6.2 percent, 4.9 percent, 4.2 percent, and 4.2 percent. The Council Report states, "The purpose of slowing down the rise in demand for 1969 to 1970 was to moderate the rise of prices. . . . The purpose is being achieved." But it does acknowledge that ". . . the process of disinflation during the year, given the degree of economic slack that has existed, has been disappointly slow" (p. 49).

The Council's program for translating a reinflation of demand into rising output rather than more rapidly rising prices relies heavily on free markets. It eschews "voluntary controls" but expresses concerns about "insulation from competitive market forces" including that due to the Federal government itself. In the latter case it acknowledges that the government "has the instruments at hand for correcting the problem," as with industries protected from foreign competition by import quotas. But it frankly recognizes the political difficulties of using the "instruments at hand."

My critical sallies should not be taken to imply that there is not a great deal of sound economics underlying the Annual Report, as we might have expected from the professional qualifications in the membership and the staff of the Council. But one may ask—and this applies to reports under Democratic as well as Republican administrations—whether we should not look for a Council which expresses the best professional knowledge that can be brought to bear, regardless of where the political costs may fall. The current Report repeatedly relates unemployment to reduced expenditures for Vietnam and the military generally. Yet, aside from acknowledging that from mid-1958 to the end of 1965 the rate of inflation was no more than 1.5 percent per year, ". . . for all practical purposes, 'reasonable price stability' " (p. 76), and noting that much of the previous post-World War II inflation related to the Korean War, the Council is singularly tactful in not charging the original inflation to the War in Southeast Asia or its stubborn persistence to the stubborn, continuing pursuit of that War. It is indeed also singularly gentle in failing to stress substantive proposals for relieving pockets of unemployment caused by reduction of defense spending.

There is a welcome acknowledgement that ". . . a permanent cut in income taxes probably has a more powerful impact than an equivalent reduction that is known to be temporary" (p. 73), a principle also applicable to the 1968 and later extended income tax surcharge intended to halt inflation. There is reference to ". . . the lagged rise of interest costs, taxes, regulated rates, and other costs or prices that joined the inflationary stream late and have kept it running" (p. 61), but no questioning of Administration (or Federal Reserve) policies which raised interest costs, taxes, regulated rates and other costs or prices. There is the conventional claim that ". . . effects on the cost and availability of financing played an important role in bringing the long boom in capital investment almost to a halt during 1970" (p. 32) despite the Council's reports of the dominant role of "strong growth trends in some industries and low operating rates in others" and the interesting appendix to the Report emphasizing that the "liquidity crisis" was ridden out very well by almost all large corporations, which account for the bulk of capital spending.

Economists may well wish to reflect on problems for which the Council Report attempts no definitive answers. What did cause the inflation and why did it persist in the face of developing unemployment? Was it just initial mismanagement or are conventional fiscal and monetary policy perhaps inadequate to cope with the kind of surge of government spending for goods and services that came with escalation of the war in Vietnam? Are there really any measures that can do much to mitigate the cost of a small but significant war which develops in the midst of a full employment economy? Particularly in the light of work by Friedman and Franco Modigliani on "permanent income," is consumption much more immune to the effects of temporary fluctuations in income tax rates than has been recognized? How much is business investment likely to be affected by changes in business taxation—either surcharges in the midst of an inflationary boom or added tax depreciation allowances in a period of substantial and growing excess capacity? Has anything occurred to make monetary policy appear more potent now, other than a change in the political winds, than it was when Keynesian views received more than lip service? "Fine-tuning" aside, what are the possibilities of bringing about intertemporal substitution of demand through fiscal and monetary policies which alter intertemporal price ratios? For example, politically unpalatable as it may be, might business investment be perhaps most usefully influenced by a high marginal "on-again, off-again" investment tax credit *and* investment tax which would make it more profitable to invest in periods of economic slack than in inflationary booms? And do not the Council's own figures and analysis demonstrate that tight money as an anti-inflationary device has a prime effect of taking the bottom out of much needed housing markets, a secondary effect of restricting urgent state and local and non-profit investment, and no effect or perhaps a perverse effect on business capital spending? Finally, is the stubborness of the inflation related not only to lags and to persistent expectational factors, which continuation of the War may well be a considerable obstacle to reversing, but also to the battery of supply-reducing and cost-increasing taxes and interest rate rises which have been part of the presumably anti-inflationary policies we have pursued?

With all the sound economics in the Council Report, much of it seems somehow tentative and cautious. One may infer that the Council does not feel able to be too aggressive about policies which the Administration is not ready to support, or which the Congress would not adopt. Is it proper to acquiesce in substantial unemployment going on well into 1972, if not further, just because, having arrived at the point of espousing a full employment budget, it seems too audacious to argue that even such a budget might well be in substantial deficit? The Council suggests in a number of places that unemployment is related to defense cutbacks, while acknowledging that a great deal of it is not reasonably so related. But after literally decades of recognition of the problem of reconversion, with specific proposals of Senator McGovern in the legislative hopper, why has the Administration not been able to come up with a program for retraining and relocating people or doing whatever else is necessary to enable us to realize the goal of prosperity and full employment with peace which has been frequently proclaimed?

Many of us believe that this economy has the capability of maintaining full employment and reasonable price stability without war. I should have liked to see a Council Report boldly charting that path, wherever the political chips might fall.

REFERENCE

M. Friedman, "A Monetary and Fiscal Framework for Economic Stability," *Amer. Econ. Rev.*, June 1948, *38*, 245–64, reprinted in his *Essays in Positive Economics*, Chicago 1953.

The 1971 Report of the President's Council of Economic Advisers: Priorities and Efficiency

By ABBA P. LERNER*

In the quarter century of the life of the Council of Economic Advisers, the interconnectedness of economic problems has caused the Council's advice to the President and to the country to spread out in a number of ways from a concentration on "maximum employment, production and purchasing power" to the "other essential considerations of national policy" referred to in the 1946 Employment Act which fathered it. The most visible of such extensions was the concern with price stability. In this year's report the Council's interest is extended further into the micro-economic fields of economic priorities and the efficient use of resources.

In chapter 3 on National Priorities, the *GNP* is broken down first into the *claims* of different parts of the body politic: federal government, state and local government, households (for personal consumption), business (for fixed investment, inventory investment and residential construction), and exports. Then they are broken down into "*functions*" or different social purposes: general government, defense, investment, (fixed business investment, inventories, and net exports), new housing, education, health, transportation and "basic necessities."

The "claims" breakdown is shown not only for selected past years but in projections for 1975 and 1976. This will be seen by some as a small step in the direction of "planning" in an ongoing convergence between capitalist and socialist economies symmetrical to the converse movement by socialist countries from completely centralized planning to projections or guesses of internally consistent scenarios that would guide the decisions of their now somewhat decentralized managers.

The projections raise some questions on the more macro-economic issue of the meaning of a "full employment budget." In the President's report to the Congress, it seems to mean that budget deficits in depressions will be kept in check by limiting government expenditures to the level contemplated in a *balanced budget* at full employment. The President's "Keynesianism" would thus seem to involve only a *partial* displacement of the pre-Keynesian principle of balancing the budget by the Keynesian principle of "balancing the economy" by keeping total spending at the full employment level; and not to include a recognition that the *maintenance* of full employment could call for a deficit or surplus.

The Council's projections for 1975 and 1976 however show budget *surpluses* of about $2\frac{1}{2}$ percent of the *GNP*. If government expenditure in depression is to be limited only by the level of tax *revenues* in such a full employment surplus budget, deficits, while still tethered, have their scope extended by the amount of the full employment surplus.

The breakdown of the *GNP* by "functions" shows that between 1955 and 1969, education, health, and general government scored the greatest increases in their relative share in the *GNP*. It is perhaps unfortunate that "basic necessities" is the name given to an item which, although it excludes education and health, amounted to 45.7 percent of the *GNP* in 1945 and 41.6 percent in 1969 (or some 60 to 70 percent of net disposable income). In view of what the world outside the United States is living on—even the rich countries—it would have been more seemly to have put a larger part of this under "All other."

---

* University of California, Berkeley.

Chapter 4 on "Economic Growth and the Efficient Use of Resources" goes more deeply into microeconomics. After some comments on growth, it turns to the current concern with environmental pollution as a peg on which to hang the basic lessons of the market mechanism. Here again, and this time more clearly, can be seen a convergence of western and eastern economics in a symmetrical turning towards prices as guides to the more efficient use of resources.

Pollution is the entering of new elements of the universe into the realm of scarcity. As clean air and pure water become scarce, the same problems arise as when labor became scarce with the expulsion from Eden and land became scarce as man became fruitful and multiplied and filled the earth. The resulting pains were alleviated by the use of *prices*, wages of labor and rent of land, to keep these resources from being used for less urgent purposes that could not pay the prices offered by the more urgent uses. But the application of pricing to the newly scarce elements calls for the adaptation of the price mechanism to deal with *public* goods where there is no proprietor to demand a price from those who consume (i.e., destroy) them.

Chapter 4 approaches these problems carefully and indirectly. It begins by pointing out the severe limitations of "unified development" in which the polluter is also the one damaged by the pollution so that he would try to weigh the damage from the pollution against the costs of abating it, seeking to achieve the *optimum degree of pollution* (although this expression is not used). Consideration is given to procedures that would lead to an approximation to this goal, and setting uniform physical standards of pollution-reduction is shown to be wasteful when the cost of meeting the standard is not uniform. The analysis then proceeds to compare the advantages of setting a *price* or *charge* per unit of pollution equal to the damage done per unit of pollution emitted, with the alternative of paying a corresponding subsidy for *not* polluting. It discusses when it might be better, instead, to decide on the *quantity* of pollution to be permitted and sell this quantity of transferable pollution rights or permits on a competitive market which would determine the price per unit.

A more explicit spelling out of the basic principles involved could have been of service here. There may be cases where it is clear that a certain pollutive activity should be prohibited altogether—the quantity fixed at zero output—and other cases where overwhelming difficulties of measuring the value of the damage done by a pollutant make it reasonable to fix some arbitrary standard of pollution reduction or some arbitrary quantity of pollution permitted. But in general to ask whether it is price or quantity that should be fixed is something like asking whether it is supply or demand that determines equilibrium output. What we seek is the intersection of two curves: the curve of the *marginal damage* from additional pollution and the curve of the marginal cost of reducing pollution. If we know both curves, we know the point of intersection which indicates both the price and the quantity so that it makes no difference which of these we say we are fixing. Nevertheless it is the *price* which is fundamental since it is the price which tells each polluter what is his optimum contribution to the optimum total output of pollution.

It should be noted that in Figure 1 the increasing marginal damage per *additional* unit of pollution rises to the right in the normal manner; but increasing marginal cost of *reducing* pollution must be measured in the opposite direction on a curve that *rises to the left.*

If the position of one of the two curves is known better than the other and the better known curve is more horizontal (on a diagram which shows the marginal cost, the marginal damage and the price vertically and the quantity of pollution horizontally), setting the *price* is more efficient (results in less error) than setting the quantity. If the better known curve is more vertical then it is more efficient to fix the *quantity* and let the market determine the price per unit.

The report points out that setting the quantity of pollution (and issuing the corresponding number of transferable pollution permits) provides the possibility for conservationists, or those hurt by the pollution, to buy up some of these permits if they think they are too cheap—too many of them having been issued. Such a market might be developed further by permitting *trade* among different pollution permits in a kind of arbitrage by conservationists or others who considered their relative prices to depart from the relative marginal damage. Thus if the charge is \$10 for emitting a ton of CO_2 and \$20 per ton of SO_2, those who felt that SO_2 was not twice but four times as harmful as CO_2 could buy up SO_2 permits at \$20 and trade them in at the government pollution permit office for twice as many CO_2 permits, which they could sell to CO_2 producers at \$20 for two. They would have improved the atmosphere at no cost to themselves except perhaps for a small fee charged by the pollution permit office.

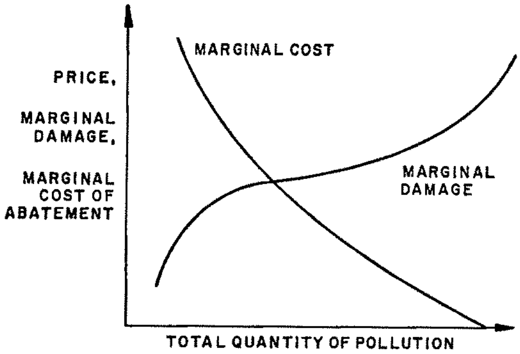


FIGURE 1

The statement is made that with prices set for pollution, "Discovery of new techniques to control pollution would be encouraged because every reduction to pollution would lower the payments for the right to emit pollutants" (p. 117). This is correct if "lower the payments" means that less would be paid when less pollution is emitted. But if—as the context seems to suggest—it means that a reduction in pollution would lead to lower charges per unit of pollutant emitted because of a diminished marginal damage, then this is not so for any "competitive" polluter who does not produce enough pollutants to be able to affect its price significantly. And where a firm is large enough to be able to influence the price of pollution by varying its own output, we would have a kind of monopolistic distortion "in reverse." To such a producer the increase in his pollution charges from producing an additional unit of pollution would be greater than the permit price. As the additional pollution raised the price per unit (as it should if there is increasing marginal damage from pollution), he would have to pay not only the higher price of the additional unit but *the price increase* on each unit that he was previously producing.

If he engaged in pollution abatement up to the point where the marginal cost of this activity is equal to the corresponding marginal revenue (the reduction in the pollution charges he has to pay), he would be going *beyond the social optimum* in his pollution abatement. This excessive pollution abatement, the "monopoly in reverse," could alternatively be expressed by saying that the marginal cost to him of producing pollution (being greater than the charge per unit) is greater than the marginal damage (equal to the charge) so that he would be producing *too little pollution.*

A failure to bring out the primacy of the *price* or *charge* for damage done as the proper application of the price mechanism also leads to some troubles in the treatment of the "pollution abatement subsidy" (p. 118) where it is argued that "in order to keep a subsidy payment down, the government agency would have to incur the expense of ascertaining the level of pollution that would be incurred without any pollution control" and with new products "no record of past pollution would be available."

But such a subsidy is nothing but the combination of a price or charge per unit of pollution together with a lump sum compensation. There is no justification for compensation in the case of new products, or even new plants in old firms producing old products, so the difficulty does not exist. Compensation is justified only where a plant was established *before* the pollution control was instituted. The polluter can then claim that the burden of the abatement should fall on the community as a whole that benefits from the abatement and that he should not be liable for more than his share of the taxes to pay the compensation. There is no justification for any claim of compensation for plants set up after the establishment of the pollution control. Those who set them up knew of the pollution charge before they made their investment.

From this exercise in working out price mechanisms for a new field where they have yet to be developed, it was a natural step to turn to a field where the price mechanism is not working properly because it has been perverted by inappropriate regulations. Outstanding among these cases are the regulations which were originally intended as protection against monopoly but were applied by administrators who have not understood the function of price as a signal of the *cost* of a service—the social damage involved in "using up" a scarce resource. They were seduced by the "value of the service" which is relevant only for the tax collector, the monopolist, or the "merchant of protection," looking for some surplus to appropriate even if it should damage rather than further the efficiency of the economy. Thus the chapter goes on to analyze, and consequently to condemn, the regulations that have grown around our transportation industry. Here the price mechanism needs only to be set free and the industry liberated from the regulations that have strangled it.

A quick look at problems connected with exhaustible natural resources results in their subjecting public utility regulation to similar criticisms, though nothing like as severely as in transportation. The chapter concludes with some consideration of health care where they point to many unsolved problems.

The extension of the interest of the Council of Economic Advisers in these chapters into microeconomics performs a valuable service in directing current hysterical reaction to past shameful neglect of pollution into sober considerations of the nature of the genuine problems involved and how they can be approached with reason and effectiveness. It will provide a welcome stimulant both to needed further research and to immediately practicable measures for increasing economic efficiency while bettering the quality of life.

The 1971 Report of the President's Council of Economic Advisers: Micro-Economic Aspects of Public Policy

By JEROME L. STEIN*

The micro-economic section of the Council Report attempts "to explore selected program and policy issues that will require new thinking if our economic system is to make its maximum contribution to national well-being." Most of the analysis is devoted to two topics: safeguarding the environment and transportation. The Council concluded its analysis of transportation by stating that a deregulated transportation industry would better serve the public interest than do the current regulations. The social cost of providing transportation services would thereby be reduced, and the economy would move closer to the utility possibility envelope. This recommendation should be supported by every economist who is not a special pleader.

Unlike the 1969 Report, written by President Johnson's Council of Economic Advisers, there is practically nothing in the 1971 Report concerning the strategies available for eliminating poverty, recent progress in this direction and the task that remains. Nothing is said in the 1971 Report about income maintenance programs or the reform of current welfare systems. This omission is the outstanding characteristic that differentiates the Nixon Council Report from those written in the Johnson and Kennedy Administrations.

The Secretary of Health, Education and Welfare wrote that:

> Health care in the United States is a current example of a vast social issue encrusted with a layer of invention and illusion. We all know there is something wrong with the current health care system, and it is commonly held that too few doctors, greedy insurance companies, and an apathetic government are at fault. But are these the real problems?[1]

Surprisingly, only two and a half pages are devoted to the subject of health care in this Report. There is neither an evaluation of alternative programs to improve health care in America, nor is there a discussion of the economics of health care.

Another conspicuous omission is the problem of revenue sharing and the plight of the cities. Although the Administration has advocated both general and special revenue sharing, there is no economic analysis of the Administration proposal in this Report. Since the major economic problems of the 1970's: health care, welfare reform, poverty and urban blight are not discussed in this Report, it is unlikely that it will arouse much interest, excitement or enthusiasm. Let us now turn to the topic which is given pride of place: safeguarding the environment.

---

* Brown University. I am indebted to George Borts, Anthony Fisher, and Mark Schupack for their perceptive criticisms of an earlier draft. Fisher and the graduate students in my course in welfare economics saved me from several errors. I am indeed grateful to them.

[1] Elliot L. Richardson, p. 37.

I. The Position of the Council

Every economist will agree with the general principles stated by the Council concerning the use of the environment.

> . . . we seek . . . a set of rules for use of the environment which balances the advantages of each activity against its costs in other activities foregone. We want to eliminate pollution only when the physical and aesthetic discomfort it creates and its damage to people and things are more costly than the value of the good things—the abundance of industrial or farm products and efficient transportation—whose production has caused the pollution. [p. 114]

The Council's economic analysis is very welcome because only the cost of pollution, and not its marginal product, attract the attention of the public. For example, President Nixon's attempt to tax the sulfur content of coal and oil, as an incentive for reducing sulfur oxide emissions, has run into heavy opposition. Officials of the Environmental Protection Agency were disturbed. They said that this taxation strategy seemed to justify the environmentalists' fears that the industry would look on the tax as "a license to pollute."[2] Economists have an educational task to perform.

To implement the above rules, we must try to answer several questions. Then the philosophical and logical disagreements among economists will surface. First, at what level of government should the rules be made? Second, how much pollution (if any) should be tolerated? Third, how should changes in this amount be determined over time? Fourth, how can we "ensure that the tolerated degree of pollution occurs for the most productive reasons"? Fifth, can we devise a set of rules and procedures which are not too costly to administer?

The Council answers these questions in the following manner.

(1) Many, though not all, pollution problems are local in character, and therefore determination of the appropriate level of environmental quality in these cases is likely to be more accurate if it is done locally rather than by the Federal Government. . . . It is not clear, however, that the Federal role should extend beyond the setting of such minimum standards where most benefits and costs of pollution are borne locally. [p. 121]

(2) The level of government should not require that each polluter reduce his pollution by the same percentage.

Once the government has decided upon the optimal level of pollution, each potential polluter should be charged a price for each unit of pollutant he discharges. Then:

> . . . he would have to compare this with the costs of pollution control in his particular circumstance. If control costs were relatively low, he would engage in extensive control to avoid paying the price being charged for polluting. If control costs were high, less control would be undertaken. Since sources with low control costs would carry out more than average control and those with high control costs less than the average, a given level of environmental quality could be attained with the expenditure of less productive resources than if all sources had to meet a common standard. [p. 117]

This charge could be imposed in the form of effluent fees: charges for emissions of pollution, or it could be done through the sale of transferable environmental usage certificates. In the latter case, the government agency would set a specific limit to the total amount of pollutants that could be emitted. It would then sell certificates granting the holder the right to emit a given quantity. These certificates could be sold at auction and would be transferable. Any institution, polluter, or group concerned with protecting the environment, could purchase these certificates. In this manner an optimum distribution of a fixed quantity of pollution would be produced.

---

[2] New York Times, April 4, 1971, p. 59.

II. A Critique of the Recommendations

A. Description of an Optimal Situation in an Isolated Locality

There are three aspects to the formulation of public policy. First, we must determine the conditions which are associated with an economy on the utility possibility envelope: the set of Pareto optimal points. Second, we must select the optimal point on the utility possibility curve in terms of a social welfare function. Third, we must decide on the public policies which will drive the economy to the optimal point. If this procedure is followed with respect to environmental policy, I think that the Council's conclusions can be questioned and also be seen to be quite inadequate to solve the problems posed. It will be proved that the price per unit of pollution damage must be uniform nationally, if the environment is considered as a national rather than as a local good. Hence, pollution control should not be left to the localities.

What conditions must exist if the economy is to be on the utility possibility envelope? These conditions imply that a given level of environment is associated with a maximum of social output, or that a given level of output is associated with a maximum level of environment.

Output y is produced by using resources, including the environment. In each locality there is a production possibility curve between output and pollution x (or its complement, the environment). To decrease pollution, abatement devices must be used, the products may be recycled and recovered, or there must be an internal process change. Each of these alternatives is costly, and requires the use of resources which could produce output. The production possibility curve is drawn on the basis of a technological and economic maximization procedure. Given the flow of output, what is the minimum level of pollution damage possible? Alternatively, given the rate of pollution damage what is the maximum output that can be obtained? This curve reflects an optimization process, as does the conventional production possibility curve.

Given the quantities of the non-environmental resources in the locality there is a positive relation between output and pollution damage: equation (1). The production function f will differ among localities. The relation between output and the minimum pollution that results will depend upon local conditions: the rate of water flow, the rate at which air circulates, etc. Pollution damage would be measured in terms of the harm created rather than by the emissions or dumping per se.

$$y = f(x) \tag{1}$$

Function $f'(x)$ is the marginal product of pollution. If the pollution rate were changed by Δx then the maximum attainable output would change by[3] $f'(x^*)\Delta x$. For example, if the firms were ordered to reduce the rate of pollution by Δx, then resources would have to be used for pollution abatement, recycle and recovery, or for internal process change, instead of for the production of output. The decline in output that would result from the abatement of Δx of pollution is $f'(x^*)\Delta x$." Production function" (1) is graphed in Figure 1.

When the environment is priced at a very low price (i.e., there are just minimal standards), then the rate of pollution will be x_2 and output will be x_2R_2. Firms will use a small amount of resources for abatement, recycle and recovery or for process change. Point R_2 is not an optimal situation in this locality because consumers value the environment relative to goods at a higher rate than do firms. The latter value the environment at $\overline{P}_2=f'(x_2)$, the slope of the production function.

---

[3] x^* is between x and $x+\Delta x$.

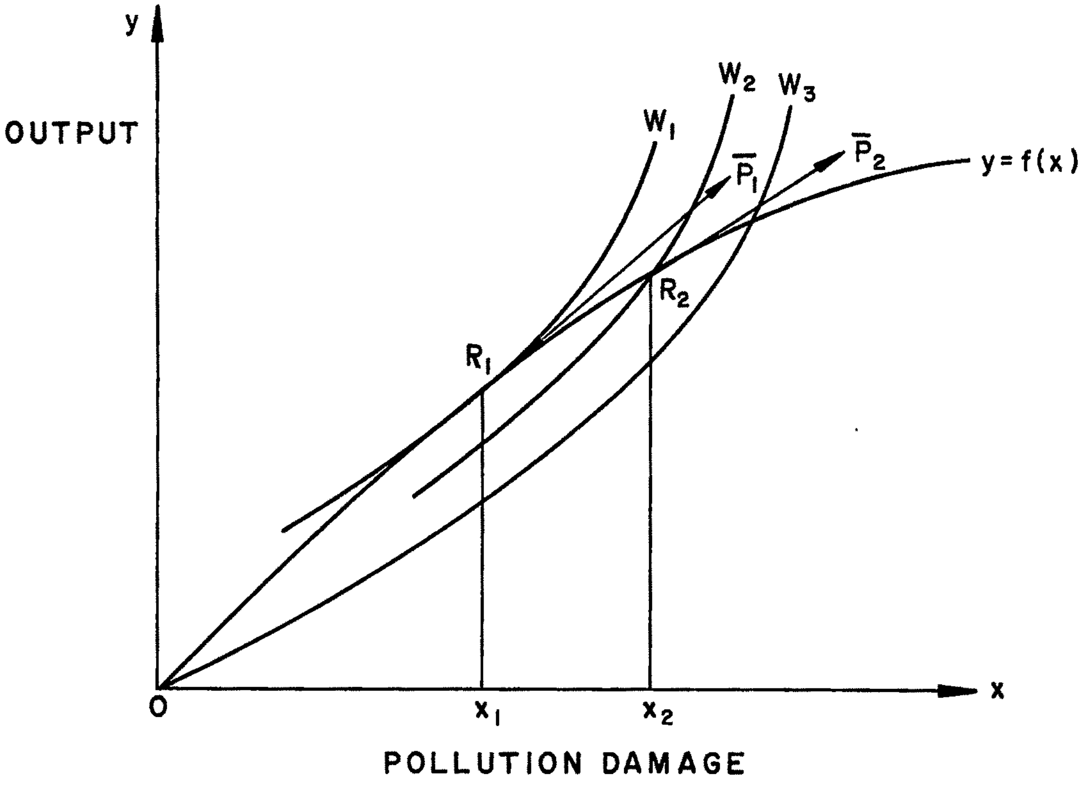


FIGURE 1

Social welfare curves W_1, W_2, W_3 have been drawn in Figure 1 on the assumption that each locality is composed of identical households which value both output and the environment. Curve W_1 is preferred to W_2 which is preferred to W_3. At point R_2 the value of the environment to households is measured by the slope of the W_2 curve. If pollution is a problem (i.e., if there is "too much" pollution), then the slope of the social welfare function at R_2 must exceed $\overline{P}_2=f'(x_2)$. If the identical households had complete control over industrial activities, then they would use x_1 of pollution to produce x_1R_1 of goods. The value of the environment would be given by $\overline{P}_1=f'(x_1)$: it would reflect both its marginal product in producing goods and the value of the environment to households. Point R_1 is the optimum output-pollution combination in this locality, if no other locality were affected.

B. Local Control is Not Pareto Optimal

The Council Report (p. 117 quoted above) suggests that it is desirable that a given level of environmental quality be attained with a minimum sacrifice of output, or that a given rate of output be attained with a minimum sacrifice of the environment. This rule is obviously sensible, but it is inconsistent with the Council policy that pollution regulation should be the prime responsibility of localities. Why?

1. *The Production Conditions are Violated*

Suppose that the marginal product of pollution dy/dx were $\overline{P}_1$ in locality I and $\overline{P}_2<\overline{P}_1$ in locality II. The reason why these marginal products differ is that there are differences in local pollution costs. Locality II imposes a lower charge upon pollution than does locality I; hence, firms in locality II use more of the environment to produce output than do firms in locality I. The higher pollution charge in locality I induces firms to devote more resources to pollution abatement than do firms in locality II. Formally, firms will use that quantity of the environment as an input in the production process such that the marginal product of pollution is equal to the price charged for its use. If $\bar{P}$ is the price charged by the locality then

$$f'(x) = \bar{P} \tag{2}$$

determines the use of the environment by firms. If $\bar{P}_1 > \bar{P}_2$, then x_1 used in locality I will be less than x_2 used in locality II.

If firms in locality II decreased pollution by one unit, then $\bar{P}_2$ of output would be sacrificed. If firms in locality I increased pollution by one unit, then $\bar{P}_1$ of additional output could be produced. The same *total* amount of pollution would exist, but output would rise by $\bar{P}_1 - \bar{P}_2 > 0$. As long as the marginal product of pollution differs between any two localities, the same level of environmental quality can be obtained with a *smaller* sacrifice of output, by shifting pollution to the locality with the higher marginal product of pollution.

If the environment belongs to the United States as a whole, then optimal policy should require that the marginal product of the environment (in terms of producing real national income) be the same everywhere. The marginal rate of transformation may be $\bar{P}_1$ in Phoenix and $\bar{P}_2 < \bar{P}_1$ in Los Angeles. If pollution damage and output were transferred from Los Angeles to Phoenix, real national output (as conventionally measured) could be constant and the level of pollution damage in the United States could be decreased as a result of such a move. Arizona loses less environment than California gains. Such a situation would not occur if the regulation of the use of the environment were the prime responsibility of the locality. The price of pollution $\bar{P}$ must be the same in all localities if we are to be on the maximal possibility curve for the United States as a whole. Underlying the concept of a maximal production possibility curve is the view that the environment does not belong to any specific locality but belongs to the United States as a whole. Hence, we wish to maximize the total environment in the United States corresponding to a given rate of national output.

2. *Consumption Efficiency Conditions will Generally be Violated*

Suppose that R_1 were the optimum point in locality I and R_2 were the optimum in locality II. The social welfare curve in locality II (not shown in Figure 1) is assumed to be tangent to the production function at point R_2. Each locality *considered as an entity* is at an optimum. Nevertheless, the society (U.S.A.) is not on its utility possibility envelope. Why?

The marginal rate of substitution between real income and the environment differs between the two localities. Households in locality I are willing to give up $\bar{P}_1$ units of output for a little more environment; but households in locality II are willing to give up some environment for only $\bar{P}_2$ of output. There is an opportunity for a profitable trade: if pollution were shifted from locality I to locality II, then both localities could be made better off. For example, suppose that locality I asked locality II if it could use the latter's environment for (say) municipal sewage disposal. It would be willing to pay a maximum of $\bar{P}_1$ for the privilege. Locality II only values the environment at $\bar{P}_2 < \bar{P}_1$. Therefore, some price $\bar{P}$ (where $\bar{P}_1 > \bar{P} > \bar{P}_2$) could be agreed upon whereby locality I uses locality II for its sewage disposal.

The argument is quite general. As long as the marginal rate of substitution $\bar{P}_1$ differs from the marginal rate of substitution $\bar{P}_2$, the United States economy is not on the utility possibility curve. Even if each locality optimized, a social optimum would not be reached unless the marginal rate of substitution $\overline{P}$ were the same in each locality.

Insofar as two localities can trade pollution for goods, as would be the case with municipal sewage, then the marginal rates of substitution will tend towards equality. However, many localities pollute each other's environment (e.g.) by using streams, rivers, and the air as waste receptacles. Therefore, they regard the adjacent communities' environment as a free good. When localities control their own environmental usage, an optimum situation is unlikely to result. It is unlikely that trade among the municipalities in the United States can lead to an equalization of the marginal rates of substitution between output and the environment among localities.

3. *Firms may Exercise Monopsony Power to Prevent Localities from Optimizing*

There is another reason why pollution regulation should not be left to the localities (who are major polluters in their own right). Suppose that the output-pollution point were R_2 in Figure 1. The town council decides that it would like to be at point R_1: it is willing to sacrifice $x_2R_2-x_1R_1$ of output for the sake of a better environment. The polluting plant says that if a unit tax $(\overline{P}_1-\overline{P}_2)$ were imposed upon its pollution, it would relocate. If the polluting plant moved out of the locality, then the level of local welfare would be reduced to $W_3<W_2$: the town would be worse off than it was with the pollution. By threatening to relocate, the polluter could prevent the locality from moving to its preferred position R_1. If the tax per unit of x were national in scope, then the threat to relocate would not be credible; and the firm would decrease pollution in the area albeit at a higher cost per unit of output.

4. *The Environment is Often a Public Good*

Suppose that the municipalities along the New England coast (acting individually) allow oil refineries to be established. They are fully cognizant of the dangers of oil spillage and air pollution, but are attracted by the income and employment potentialities resulting from the refineries. Formally, the municipality optimizes at point R_1 in Figure 1. At the town meetings, the identical households decide that the additional income to be generated by the refineries justifies the damage to the environment and the loss of tourist revenues. Does the condition of local autonomy, which is advocated by the Council, lead to a Pareto optimum?

The enjoyment of an environment is often a public good. My enjoyment of the natural beauty of Cape Cod does not reduce the enjoyment that you can derive from it (until the highways become congested). The social marginal rate of substitution between such an environment and goods is the *sum* of goods which everyone is willing to give up to protect the environment. It will exceed the marginal rate of substitution exhibited by the residents of the locality, because many people may enjoy the environment without paying its full value to the local merchants. The appropriate social marginal rate of substitution is the maximum sum in taxes that the voters of the *United States* are willing to impose upon themselves to avoid the deterioration of the New England coastline. The Council failed to consider the environment of the United States as a public good. It seems to have missed many of the subtleties of welfare economics.

On the basis of my analysis, I arrive at a conclusion quite different from that reached by the Council. A *national* pollution policy must be followed. If pollution regulation were the prime responsibility of localities (who also pollute each other's environment), then our economic system will not be making "its maximum contribution to economic welfare."

If there were a national price per unit of pollution damage, then the marginal rate of transformation between environment and output would be the same everywhere. A given level of national output would be obtained with a minimum level of national pollution damage. Insofar as production functions (f_i) differ among localities, there will be differences among localities in the amounts of pollution. For example, if $\bar{P}$ is the national price per unit of pollution, then:

$$\bar{P} = f_1'(x_1) = f_2'(x_2) = \ldots = f_n'(x_n) \tag{3}$$

among the n localities. But, there is no reason why x_i should be equal to x_j or that $(y/x)_i$ should equal $(y/x)_j$. People who like environment relative to output will move to areas where x is low, and those who like output relative to environment will move to areas where y is high.

5. *The Total Amount of Pollution*

The Council Report contains little about the optimum total amount of pollution. It simply states that the total quantity should be left to the locality. The council recommends that the political unit sell environmental usage certificates at auction; and that these certificates be transferable within its jurisdiction. To be sure, environmentalists may purchase some of these certificates. But, because there are external effects of the environment, too little would be bought by these groups.

Our previous analysis implies that transferable environmental usage certificates be sold by the federal, not by the state and local, government. Then a national price would be established per unit of pollution damage. Consumers are then confronted with a trade off, between output and the environment, in the form of a tax on the polluting product. As long as consumers as a whole wish to have more environment, they must agree to decrease the total quantity of environmental usage certificates sold, i.e., decrease their ability to purchase the polluting product. Then they will be made to realize that safeguarding the environment involves the use of resources or that pollution imposes harm upon others. The discussion concerning the quantity and quality of the environment then becomes the one of agreeing on the quantity of certificates. Polluters and environmentalists can try to influence Congress concerning the level of the national tax per unit of pollution x, or the total quantity of certificates. Given the uniform national price, the quantity of pollution will be determined by: (a) The possibilities and costs of abatement devices; (b) the existence of alternative plant and dumping sites; and (c) the substitution of nonpolluting for polluting products.

The difficult aspect of economics is not the statement of general marginal conditions but in finding methods to implement these conditions. It is not clear that the Council's recommendations serve to implement its principles.

REFERENCES

E. L. **Richardson**, "Myth and Reality: Problems of Health Care," *New York Times*, Apr. 2, 1971, p. 37.

New York Times, Apr. 4, 1971, p. 59.

The Duopoly Game: Output Variations

By D. K. OSBORNE*

Game theory has produced no important results in the field of oligopoly. It has given us a few insights, particularly the extremely important idea that firms act in strategic settings, but no concrete answers to the fundamental questions. This is not the fault of game theory; it is the fault of its users. After all, game theory is merely a way of thinking about the *general* features of games—or those conflict situations which may be viewed as games—and is necessarily too abstract to be informative about oligopoly behavior. It must be specialized—given economic content—before it will tell us anything useful about oligopoly price or production. The required degree of specialization has largely been avoided by those oligopoly theorists who use the game theoretic framework, possibly because it entails great risks (something might be overlooked).

To be sure, some of the features of oligopoly games may be specified with little risk, or at least no greater risk than is common to most parts of economic theory. Thus in many contexts the oligopoly game may be taken as one of perfect information. This is especially true of price games, where buyers provide the information, and almost as true of output games, where trade associations, common suppliers, buyers, etc., provide the information.[1] Secondly, moves may be regarded as the selection of prices, output rates, advertising expenditures, whatever seems appropriate in the context; and there is no obvious reason why all types of moves should be studied in any particular game. A price or output game would seem worthy of study in itself, in ignorance of advertising and other moves. Thirdly, though Edith Penrose, William Baumol, Robin Marris, and others have taught us to consider other things in addition to profits as managerial goals, the other things (revenues, growth, etc.) might on some occasions be ignored to our benefit. In fact, preliminary results from the more general study, of which the present paper is a report, indicate a strong strategic invariance under many changes in goals. Fourthly, the problem of side payments and characteristic functions may usefully be disregarded in nearly all contexts other than that of the cartel. Most oligopoly situations should probably be viewed in the terms of a noncooperative game with unrestricted communication.[2] At any rate, that is the framework of the present paper.

Three specifications, however, raise difficult conceptual problems: the concept of solution, the character of the strategy sets, and the matter of the length of the game. Of these three problems, that of the solution concept, curiously, has received much the greatest attention. Between the joint maximum, the Nash equilibrium point, the Paretian optimum, and that center of minor controversy, the maximin, and others as well,[3] there is no lack of contenders. If there is a dominant opinion about solutions it seems to support the joint maximum as solution to variable sum, and the maximin as solution to constant sum, games. But what a particular solution concept implies about price or output, about the welfare loss caused by oligopoly, about the size distribution of firms—of these questions we know nothing. Game theory will shed almost no light on these questions until it is infused with definite specifications of the strategy set.[4]

---

* Associate professor of economics and management, State University of New York at Albany. I am grateful to the referees for several suggestions which improved my exposition.

[1] George Stigler has argued that potential gains from secret price cuts sometimes exceed the risked losses arising from their discovery by rivals. He shows that such gains in duopoly are ordinarily negligible, but, depending on buyer characteristics, nevertheless calculable. This result can be saved in duopoly price games by appropriate modifications of the assumption of perfect information, but it would seem to have little bearing on output games.

[2] John Harsanyi is a leading student of this type of game.

What is more curious still is that such attention as has been devoted to strategies is largely confused. In fact, most discussions found under the heading of "strategies" are really about *moves*. Thus Shubik: ". . . the strategy of each player consists of naming a price at which he will sell his product and selecting a production rate or initial stock position" (p. 111).[5] And Robert Dorfman, Paul Samuelson, and Robert Solow, in those cases where they specify strategies, call them "100" and "200" (p. 424), "small, medium, and large" (p. 426), and the like. In no case can these strategies be considered, in the words of John von Neumann and Oskar Morgenstern, ". . . a complete plan: a plan which specifies what choice the player will make in every possible situation, for every possible actual information which he may possess at that moment" (p. 79). These are far from being isolated instances; indeed, I know of only one exception to the tendency to call moves strategies, and that appears in a paper published in 1954 by Alexander Henderson. He took as strategies three price policies proposed long ago as theories of duopoly: (i) the Bertrand policy of ignoring possible price reactions by the rival, (ii) the Stackelberg policy of anticipating unthinking price reactions by the rival, and (iii) a policy of trying to drive the rival out of the market, if it is not too costly to do so. As far as I can determine, this idea has not been developed. Henderson's early death cut short his exploration of the initial conception, and no one has followed his lead.

The present paper makes that attempt. The idea is to consider five explicit strategies, examine them for dominance relations, determine a set of solutions, and draw the implied conclusions about resource allocation under duopoly. The conception is attractively simple, but the analysis is vexingly difficult. Moreover, the answers it gives depend on whether the firms choose price, letting the buyers decide how much to buy, or outputs, letting the buyers decide how much to pay. That is, *duopoly price games are not strategically equivalent to output games*. I deal here with the output game, partly for reasons of relative computational and expositional simplicity, (the output game being simpler than the price game) partly for reasons of historical continuity (the strategies analyzed here having been proposed in the form of output variation) and partly because one begins somewhere.

Whether duopolists are, in practice,

---

[3] See Daniel Ellsberg, Harsanyi, Anthony Koo, John Nash, Hans Neisser, and Martin Shubik. The best general survey of solution concepts is found in R. Duncan Luce and Howard Raiffa.

[4] Game-theoretic discussions of solution concepts are paralleled in oligopoly theory per se, where the main contenders are the joint maximum and the proposals of Cournot and Stackelberg. The joint maximum apparently enjoys the greatest professional support (see William Fellner and Stigler) if only when cast in a suitable vague form, such as Fellner's, pp. 33, 134. But viewed in the light of game theory, these discussions are concerned with strategies, too; it is just that the strategy set contains one member at a time. Because it is explicit, this strategy says something about resource allocation under oligopoly. That is why the classical theories, naive as they are, have not disappeared, and why game theory, sophisticated as it is, has had so little impact.

[5] In his most general game, Shubik has the moves "consist of actions involving production, financing, payment of dividends, advertising, research, expansion, and other actions of a firm" (p. 215). Clearly, everything is covered!

price or quantity adjusters is not as clear as might be supposed. It is true that they name prices, but they must also set up production schedules. Conceivably, they regard output as their action variables, and in naming prices only fulfill the role of Walras' auctioneer. But even if one believes, as I do, that price is the usual action variable, he should not think that the study of output games is an idle exercise. Not only is it useful and interesting to look at old propositions in a new light; it is intellectually pleasing to know which classical duopoly theories represent maximin solutions, which are Nash equilibrium strategies, and so on.

There are two strands to the analysis. In the first, I am regarding the classical duopoly theories as serious candidates for the role of *equilibrium* strategies followed by actual firms, and with their aid seeking the allocative effects of duopoly. In the secondary strand I am investigating the game theoretic properties of the classical strategies (and also of a more recently proposed strategy, that of Robert Bishop, which is too good to ignore).

The main results are, firstly: the welfare effects of price-taking duopoly are in any case bad; but they can be more or less bad depending on the relative costs and relative market shares of the firms, being worse when the firms are exactly alike in these two respects and less serious when the relation between the firms' costs is much different from the relation between their market shares. This finding that similarity of duopolists' positions is bad for consumers is not a new one, as adherents of the joint maximum theory of duopoly will quickly point out; but the result is found here without the aid of the joint maximum solution, which completely lacks rationalization. Secondly, of the classical duopoly solutions, both Stackelberg points and one or the other of the "market share" points (where one firm holds on to its current market share and the other correctly foresees that it will) are Nash equilibrium points. More interestingly, the Cournot solution is usually maximin: if both firms wish to follow the strategy whose worst outcome is better than the worst outcomes to their other strategies, they will usually behave as Cournot assumed.

I. General Features of the Duopoly Game

Under this heading stand the state of information, the characteristics of market demand, the cost structures and objectives of the firms, and the rules of cooperation as defined by the legal environment of business. Detailed specifications in this regard suffice, when adjoined to the rules concerning moves and to a theory of solution, to completely determine the game. These specifications cannot be accomplished in an a priori obvious way; they are informed by one's point of view, by one's theory of business behavior. The specifications set forth here form my own point of view, which the reader may judge according to his taste. Rather than engaging in an elaborate and possibly unconvincing justification, I will merely list and comment briefly on the main features.

Information. In the technical sense, information is assumed perfect: each firm knows precisely how much the rival firm is producing and has records of both its own and rival's past production rates.

Firms would find it very difficult to keep their production rates secret, what with the information-spreading role of suppliers, buyers, company reports, goings on in factor markets, and the like, but it is my view that firms do not even desire secrecy with respect to prices and production rates. The great over-riding concern of duopolists is the attainment of an *entente* with respect to prices and production, and this cannot be attained without open knowledge.[6] Secrecy is desired, and assiduously pursued, when it is a matter of new product developments, new production processes, advertising campaigns: in short, those activities which cannot *quickly* be matched or nullified by rivals. These receive the main force of the competitive drives of enterprise under most modern conditions, whether their locus is duopoly or not. "Competition" in the economist's sense is only a dissipation of energies to no good account.

This is not to say that firms are indifferent to the features of a possible *entente*, for the conflict of interests never disappears; it is merely to say that the *entente* is so necessary that it is cultivated by the flow of information on prices and production. Firms often make public announcements of their views about the appropriateness of current prices and production, and while these are doubtless thought to perform some useful public relations function, they also serve to remove unsettling suspicions and expectations from the minds of rivals.[7]

Another sense in which information is perfect is the usual one concerning goals, costs, and market demand. Both firms know their rival's utility functions and their own and rival's cost functions—or have the same estimates of them—and identically perceive market demand.[8] This assumption cannot be so strongly supported as the previous one, for differences of opinion about market demand are possibly a major cause of friction in duopolies. However, such differences cannot be great when the market is in equilibrium. Since the present study is limited to the equilibrium properties of the duopoly game (it is concerned with the game in normal form) the assumption of perfect information in this sense seems the appropriate one, even though it does away with much interesting behavior.

Cost and Demand. The market price p is a linear function of quantities x_1 and x_2 offered for sale: $p=D-d(x_1+x_2)$. This homogeneous product demand function may, by appropriate choice of commodity unit, be written no less generally as

$$p = D - x_1 - x_2 \tag{1}$$

Production proceeds at constant marginal cost. Ignoring fixed costs, which play no role in the present analysis, we may write the cost functions as

$$c_1 = ax_1(0 \leq a \leq 1) \tag{2}$$

$$c_2 = (1 - a)x_2 \tag{3}$$

By varying a over the closed unit interval, we may investigate the effects of cost inequality on the strategic properties of the game. Moreover by varying the strength of demand, D, we may study the effect of market size on the strategic properties; and it may be shown that increasing D yields the same effect as lowering the upper bound on a.[9]

---

[6] Once a tenable position has been reached with the aid of perfect information, each firm might be tempted to depart unilaterally from it. I would prefer to leave a study of the consequences to an analysis of the Stigler type. This will be harmless if the opportunities for secret evasion of an agreement do not condition the nature of the agreement.

[7] For obvious reasons, these occur most often in homogeneous oligopolies. Steel and aluminum are cases in point. Price changes are typically announced weeks ahead of their effective dates, to reduce the chances of rival misunderstandings. Of course, in these industries list prices serve merely as a basis of agreement which the firms attempt to evade. This evasion, however, is not always—perhaps not ever—secret.

[8] Having perfect information, the firms need no mixed strategies. We therefore need not consider cardinal utility, which is only required to give meaning to mixed strategies. All the solution concepts employed are definable in terms of greater or less, i.e., in terms of ordinal utilities.

The firms' knowledge of each other's goals is necessary if we are not to put further restrictions on the utility functions. See John Kemeny and Gerald Thompson.

[9] Equations (2) and (3) are not the most general

Managerial Objectives. Both firms, whose utility functions are ordinal, seek maximum profits. In an important sense, this assumption is less restrictive than it might seem, for there are no discontinuous changes in the strategic properties of the game when a is put equal to zero or one (equivalent to one or the other firm maximizing revenue) or as D approaches infinity (equivalent to *both* firms maximizing revenue, for the average variable costs will be so small in relation to D as to be negligible).

The utility functions could be generalized without great conceptual cost, but the computational cost would be serious. However that may be, the objectives can be written

$$\text{for firm 1: } \max P_1 \equiv \max_{x_1 \geq 0} \left[(D - a)x_1 - x_1^2 - x_1x_2\right] \tag{4}$$

$$\text{for firm 2: } \max P_2 \equiv \max_{x_2 \geq 0} \left[(D + a - 1)x_2 - x_2^2 - x_1x_2\right] \tag{5}$$

Communication and Cooperation. The firms pursue their objectives with the aid of unrestricted communication. Like those in the real world, they communicate in face-to-face meetings at trade conferences and the like, by telephone or letter, and, if necessary, through the public press. *No binding agreements or side payments, however, are permitted.* The firms may cooperate, if they are not too obvious about it, but their interests remain separate, and such agreements and understandings as they reach are not legally enforceable. They may develop the most elaborate market sharing arrangements, such as were shown in the "Electric Companies Case," though such complicated formulas are hardly necessary in duopoly; but closed to them are side payments, revenue pooling, and the like. These devices are only important in cartels, which require a separate analysis, so this theoretical simplification seems empirically appropriate.[10]

Solution Concept. I define the solution to this game as those members of the set of Nash equilibrium points which are also Pareto optimal from the standpoint of the firms, ignoring the interests of consumers. Because the firms choose output under conditions of perfect information, they will be able to select strategies that are "best responses" to opposing strategies; because they are prevented from pooling their individual gains for subsequent division, they can do no better than select a best response strategy. Hence the equilibrium point. On the other hand, because the firms may communicate as necessary and coordinate their actions as desired, they may be regarded as a coalition playing a two-person game against consumers. Since they are stuck with their individual gains from equilibrium strategies, however, both will prefer an equilibrium point A if it yields *both* firms greater profits than an equilibrium point B. Moreover, if A yields greater profits to firm one, but equal profits to firm two, the latter will still pre-

---

linear cost functions (even ignoring fixed costs). Suppose we "start" with the functions $p=D-d(x_1+x_2)$, $c_1=b_1x_1$, $c_2=b_2x_2$, and $b_1+b_2\neq d$; then we cannot derive equations (1) through (3) of the text. Thus if we wish to simplify notation by choosing units we can put D and either b_1, b_2, or d equal to unity; e.g., $p=1-x_1-x_2$, $c_1=b_1x_1$, $c_2=b_2x_2$. This notation does allow symmetrical expressions of the firms' strategies and profits, but it makes an investigation of the effects of the strength of demand notationally cumbersome, since it requires comparing b_1+b_2 with unity. Moreover, certain important inequalities derived below become more complicated. These considerations, however, would not be decisive if any important results depended on the relation of d to b_1+b_2. The set of good strategies depends on b_1 relative to b_2, and on b_1+b_2 relative to D, but not on b_1+b_2 relative to d. Thus the exposition is simplified with no great loss of generality if it is assumed $b_1+b_2=d$.

[10] In one sense, this is the opposite of a theoretical simplification, for it removes all justification for the joint maximum solution. But the sort of "simplification" which ignored the illegality and difficult concealment of side payments would, to use Stigler's charge against limit-pricing theory, solve the oligopoly problem by murder.

fer A to B because it needs the former's good will. Hence the Paretian character of the solution.[11]

II. Strategies

A firm may select from an infinite number of strategies; the reader will understand if I do not consider them all. Instead, I am going to pour old wine into new bottles. As I have stated above, my aims are two. One is to find a manageable framework for the study of oligopoly; the second is to explore the game-theoretic properties of the classical theories of duopoly. Even though the first purpose takes precedence, I am far from being apologetic about a certain *deja vu* flavor in the analysis, for I believe that all the classical strategies enjoy strong a priori support as possible plans of action by the firm.

Cournot. According to Cournot, the firm selects its rate of output on the assumption that its rival will keep its current rate of output fixed whatever the firm does. This assumption then yields a definite rate of output for one firm as a function of the rival's output. Now clearly, this is a strategy, a complete plan of action. It is undoubtedly a very simple strategy, and requires a degree of blindness of its user. Possibly, however, such rival reactions as occur do so with a time lag, and are mixed up with changes in costs and market demand to the extent that they are not certainly observable. But in any case, *this strategy is not dominated* by any of the other strategies considered here—an interesting result from the standpoint of the history of theory.

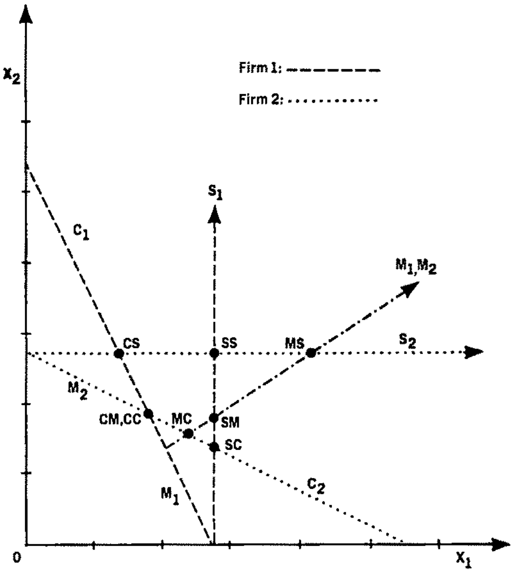


FIGURE 1. C, M, AND S LOCI ($a=1/2$, $D=6$, $k=2/3$)

Maximizing P_i on the Cournot assumption yields the reaction functions, or as called here, the *Cournot loci* C_i, with equations

$$(6) \qquad x_i(C) = (D - a_i - x_j)/2,$$

for i and j equal 1 or 2, and $i \neq j$. Of course, $a_1 + a_2 = 1$. These loci intersect at the *Cournot point CC*, in Figure 1.

Stackelberg. The Stackelberg strategy is also a simple one, but not so simple as the Cournot. In fact, it is posited on the expectation that the rival will play the Cournot, and is designed to take maximum advantage of the rival's choice: it is the "best response" to a Cournot strategy. Consequently, it too is undominated in the present set.

Substituting $x_j(C)$ for x_j in firm i's profit function, and maximizing with respect to x_i, yields the *Stackelberg loci* S_i, with equations

$$(7) \qquad x_i(S) = (D - 3a_i + 1)/2$$

S_1 intersects C_2 at the first firm's *Stackelberg point SC*; the second firm's Stackelberg point is CS. The loci S_1 and S_2 intersect at the *Stackelberg disequilibrium point SS*. See Figure 1.

---

[11] Harsanyi presents arguments supporting a further reduction of the set of equilibrium points, obtaining a solution much like the of the "ψ dominance" type of Luce and Raiffa. The arguments are not without intuitive appeal, but they naturally lack the force of those supporting merely the Pareto optimal equilibrium points.

Market Share. Another possible plan of action is to hold on to the current market share. Observe that market share is not being sought for itself, but as a possible way to maintain or increase long-run profits. Always, the objective is given by (4) or (5); maintaining current market share is simply a possible way of achieving this. Now it turns out that the strategy of always holding on to current market share, come what may, is dominated by a closely related strategy—that of holding on to current market share unless a larger share is more profitable.[12] Specifically, the *Market-share strategy* is one of keeping x_i in its current relation to x_j unless $x_i(C)$ yields a larger market share as well as greater profits.[13] Define

$$k = x_2/x_1; \tag{8}$$

k is the slope of a radius vector in output space, where x_2 is measured vertically—i.e., the slope of a *k-radius*. Then, the *Market-share* loci M_i are given by the equations

$$x_1(M) = \max[x_1(C),\, x_2/k], \tag{9}$$

$$x_2(M) = \max[x_2(C),\, kx_1] \tag{10}$$

Diagrammatically, the M_i locus 'starts' at firm i's monopoly point (where $x_j=0$ on the C_i locus) follows C_i to its intersection with the k-radius, then follows the latter outwards[14] (see Figure 1). The M_i loci intersect the other loci at points MC, MS, CM, SM, and MM, where, as always, the first firm's strategy is denoted by the first letter of an intersection point label. MM, of course, is not determinate; it could lie anywhere on the coincident parts of the M_1 and M_2 loci.

Nichol. This is the best response to the Market-share strategy.[15] Firm i substitutes $x_j(M)$ for x_j in its profit function and maximizes with respect to x_i. The result is a bit complicated.

Let

$$x_1^0 \equiv (D + a - 1)/(2k + 1) \tag{11}$$

denote the x_1 coordinate of the corner of the M_2 locus; i.e., the x_1 that equates $x_2(C)$ and kx_1. Then in playing M_2 the second firm chooses

$$x_2 \begin{cases} = x_2(C) & \text{if } x_1 \le x_1^0, \\ = kx_1 & \text{if } x_1 \ge x_1^0 \end{cases}$$

To play its Nichol strategy N_1, therefore, firm one maximizes

$$P_1(NM) \begin{cases} = f(x_1) \equiv x_1[D - a - x_1 - x_2(C)] \\ \qquad\qquad\qquad\qquad \text{if } x_1 \le x_1^0, \\ = g(x_1) \equiv x_1(D - a - x_1 - kx_1) \\ \qquad\qquad\qquad\qquad \text{if } x_1 \ge x_1^0 \end{cases}$$

We already know that $x_1(S)$ maximizes f; there are two cases to consider:

(i) $x_1(S) \le x_1^0$,

and

(ii) $x_1(S) > x_1^0$

The condition for (i) holding is $k \le q$, where[16]

$$\begin{aligned} q &\equiv x_2(SC)/x_1(SC) \\ &= (D + 5a - 3)/2(D - 3a + 1) \end{aligned} \tag{12}$$

Hence when $k \le q$, firm one's Nichol output is its Stackelberg output.

---

[12] This resembles a proposal of Franklin Fisher's concerning firms' *goals*—that firms seek maximum profit subject to no loss of market share. In the present analysis it is only a *policy* which might achieve the goal of maximum profits.

[13] By the "current" relation between x_1 and x_2, I mean the one obtaining at any stage of play, not merely at the beginning.

[14] This is only one of many possible Market-share strategies. Some non-current market share could be sought, and the profit maximizing rate of output could be conditional on some non-Cournot assumption.

[15] Named for A. J. Nichol, who I think was the first to consider this type of policy.

[16] Equation (12) follows from the simultaneous solution of equations (6) and (7), putting $a_1=a$, $a_2=1-a$.

If (ii) holds, there are two further cases to consider. Let

$$(13)\qquad x_1^g \equiv (D-a)/2(k+1)$$

denote the maximizer of g. The two further cases are

$$(\text{iia})\qquad x_1^g \leq x_1^0,$$

and

$$(\text{iib})\qquad x_1^g > x_1^0$$

The first holds when $q<k$ and $a \geq 1/2$, and also when, a being less than $1/2$,

$$(14)\quad q < k < q_1 \equiv (D+3a-2)/2(1-2a)$$

Consequently, (iia) implies that x_1^0 maximizes $P_1(NM)$. But if (iib) holds, x_1^g maximizes $P_1(NM)$.

In summary, then, the first firm's *Nichol locus* is given by

$$(15)\qquad x_1(N)\begin{cases} = x_1(S) & \text{if } k \leq q; \\ = x_1^0 & \text{if } q < k \leq q_1 \\ & \qquad \text{and } a < 1/2, \\ & \text{or if } q < k \\ & \qquad \text{and } a \geq 1/2; \\ = x_1^g & \text{otherwise.} \end{cases}$$

(It is assumed here that $D>2$, which is necessary anyway if there are to be positive Stackelberg outputs for all a in the closed unit interval.)

Now the same line of reasoning applied to the second firm's Nichol locus yields

$$(16)\qquad x_2(N)\begin{cases} = x_2^g & \text{if } k < s_1 \\ & \qquad \text{and } a > 1/2; \\ = x_2^0 & \text{if } s_1 \leq k < s \\ & \qquad \text{and } a > 1/2 \\ & \text{or if } k < s \\ & \qquad \text{and } a \leq 1/2; \\ = x_2(S) & \text{if } s \leq k, \end{cases}$$

where

$$(17)\qquad x_2^g \equiv k(D+a-1)/(k+1),$$

$$(18)\qquad x_2^0 \equiv k(D-a)/(k+2),$$

$$(19)\qquad s_1 \equiv 2(2a-1)/(D-3a+1),$$

$$(20)\qquad s \equiv x_2(CS)/x_1(CS) = 2(D+3a-2)/(D-5a+2)$$

The Nichol loci intersect others at points which have "N" in their labels. If $k \leq q$ the locus N_1 is coincident with S_1, and if $k \geq s$ the locus N_2 is coincident with S_2. If the second form of equation (15) holds, the points NM and NC are coincident, while if the second form of (16) holds, CN and MN are coincident.

Bishop. The final and most recent bit of old wine is a variant of Robert Bishop's proposal. In its broad features, the Bishop strategy is one of accommodating self-restraint, provided the rival is at least as accommodating, but one of defensive maintenance of a "proper," or "minimally tolerable" position in the market if the rival becomes too selfish. As employed by firm one, the strategy could take the following form: The firm accepts any outcome which lies on "its end" of the contract curve[17] (i.e., the end closest to the x_1 axis) and will allow itself to be moved toward the rival's end, but only so far as its market share remains at or above a minimally tolerable level. This is the *conciliatory phase* of the strategy. If the rival produces at a rate greater than that associated with the firm's minimally acceptable share—if he tries to take advantage of the firm's gentlemanly conduct—the firm will cease its conciliatory behavior and will maintain its minimally acceptable market share;

---

[17] The contract curve is the locus of points (x_1, x_2) where $P_i(x_1, x_2)$ is maximum for constant $P_j(x_1, x_2)$. In Figure 1 it would be a straight line joining the two monopoly points (the points $(x_1(C), 0)$ on C_1 and $(0, x_2(C))$ on C_2). But if $a \neq 1/2$, it would be a concave curve connecting the monopoly points.

it passes to the *defensive phase* of the Bishop strategy. In response to further increases in x_2 it will continue in this phase until its profits reach the vanishing point. Thenceforward, instead of maintaining the market share, the firm will simply maintain its rate of output, letting the second firm take the greater losses if it should further increase output. This is the *third phase* of the Bishop strategy.

The second firm's Bishop strategy is similarly conceived. Clearly, its minimally acceptable market share need not be compatible with the first's. That is, the two parts of the contract curve corresponding to the two conciliatory phases of the strategies may not meet: neither firm is quite conciliatory enough, and as a result their Bishop strategies intersect in at least one of the third phases. On the other hand, the two parts of the contract curve might overlap: each firm is so eager to achieve accommodation it is willing to "give up" more than necessary. Here the Bishop intersection could be any point on the overlapping parts of the conciliatory phases.

A very large number of Bishop strategies could be considered, one for each choice of a "reasonable" minimally acceptable market share by each firm. Bishop has investigated some of these possibilities. However, for our present purpose I propose the simplest course: both minimally acceptable shares are found at the same point on the contract curve—namely, where the curve is crossed by the radius through SS. This may or may not appear plausible but in its defense one can observe that, of the C and S strategies, a choice of S will give a firm the largest market share it can guarantee by its own actions. (If firm one chose C, the best share it could expect would be that at CC; it could, however, end up at CS, with a smaller share. If it chose S, the smallest share would be that at SS, which exceeds that at CS; it might even end up at SC, where of the points CC, CS, SC, SS, its share is greatest. The same applies, *mutatis mutandis*, to the second firm's choices.) The choice is entirely in keeping with the conciliatory nature of the Bishop strategy: the firm does not insist on too great a share to make accommodation impossible. It also preserves the self-interest inherent in the situation: the firm insists on as large a market share as it could gain from the extremely uncoordinated outcome of Stackelberg disequilibrium.

Define

$$r \equiv x_2(SS)/x_1(SS) = (D + 3a - 2)/(D - 3a + 1) \tag{21}$$

Then (see Figure 2) firm one's Bishop locus runs from its monopoly point along the (in this case linear) contract curve to the point BB, thence along the r-radius to the point SS (it happens that for the data of Figure 2, P_1 vanishes at SS) and thence vertically. Firm two's Bishop locus runs from its monopoly point to BB, thence along the r-radius to SS, thence horizontally.

As Bishop proposed it, his strategy involved the contract curve, which is linear

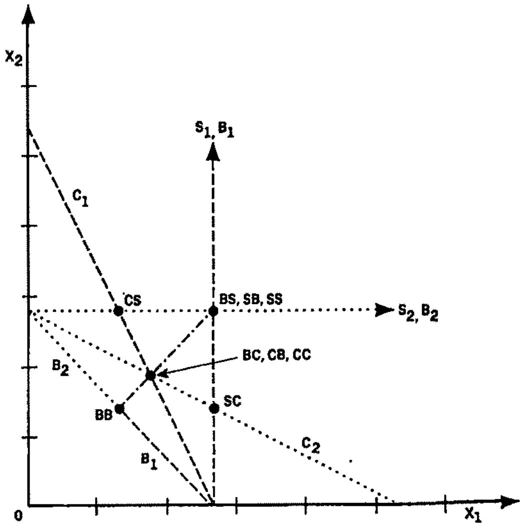


FIGURE 2. B, C, AND S LOCI ($a=1/2$, $D=6$)

$$
(22) \qquad x_1(B) \begin{cases} = (1/2)(D-a) - x_2 & \text{if } x_2 \leq (D-a)(D+3a-2)/2(2D-1), \\ = x_2/r & \text{if } \dfrac{(D-a)(D+3a-2)}{2(2D-1)} < x_2 \leq \dfrac{(D-a)(D+3a-2)}{2D-1}, \\ = (D-a)(D-3a+1)/(2D-1) & \text{otherwise} \end{cases}
$$

when the variable costs are equal, but not otherwise. In order to simplify the computations I shall modify the strategy to this extent: instead of the conciliatory phase of the strategy being coincident with part of the contract curve, it is defined by the *conciliation locus*. For firm i, this is defined simply as a straight line from firm i's monopoly point to its point of maximum profit on the r-radius. If variable costs are equal ($a=1/2$) the two conciliation loci cover precisely the contract curve, but if the costs are not equal the loci do not meet. In particular, the low-cost firm's conciliation locus joins the r-radius to the right of the corresponding join of the high-cost firm's. (See Figure 3, where, additionally, the third phases of the Bishop loci occur at different points; this is always the case when $a \neq 1/2$.)

Although the Bishop loci are coincident in more than one point, the profits of both firms vary in the same direction along the coincident parts of their loci, and we can therefore designate the *Bishop point BB* as that which is the *best* attainable by simultaneous pursuit of Bishop strategies. With both firms following this strategy, surely at least one of them will bring the outcome to a preferred point.[18]

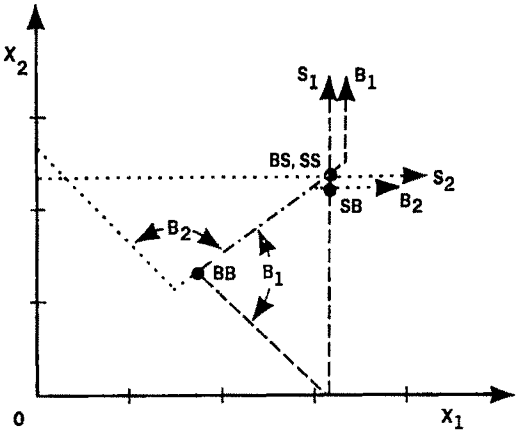


FIGURE 3. *B* AND *S* LOCI ($a=1/4$, $D=6$)

To find firm one's conciliation locus, maximize P_1 subject to $x_2=rx_1$. From equations (4) and (21) it is seen that the point

$$x_1 = (D-a)(D-3a+1)/2(2D-1),$$
$$x_2 = (D-a)(D+3a-2)/2(2D-1)$$

is firm one's best on the r-radius. Hence the line joining this point to its monopoly point is firm one's conciliation locus; its equation is

$$x_1 = (1/2)(D-a) - x_2, \quad \text{for } 0 \leq x_2 \leq (D-a)(D+3a-2)/2(2D-1)$$

That gives the conciliatory phase of the B_1 strategy. The defensive phase, of course, is $x_1 = x_2/r$, provided $P_1 \geq 0$. The third phase then begins where $P_1=0$, subject to $x_2=rx_1$; i.e., at the point

$$x_1 = (D-a)(D-3a+1)/(2D-1),$$
$$x_2 = (D-a)(D+3a-2)/(2D-1)$$

Accordingly, firm one's Bishop strategy is given by equation (22).

The same argument, applied *mutatis*

---

[18] Bishop has considered the ways in which firms could make their Bishop strategies known, even in the absence of verbal communication. These dynamic problems will not concern us.

$$(23)\quad x_2(B)\begin{cases} = (1/2)(D + a - 1) - x_1 & \text{if } x_1 \leq (D + a - 1)(D - 3a + 1)/2(2D - 1), \\ = rx_1 & \text{if } \dfrac{(D + a - 1)(D - 3a + 1)}{2(2D - 1)} < x_1 \leq \dfrac{(D + a - 1)(D - 3a + 1)}{2D - 1}, \\ = (D + a - 1)(D + 3a - 2)/(2D - 1) & \text{otherwise} \end{cases}$$

---

mutandis, yields firm two's Bishop strategy as shown in equation (23). The entire set of strategies is diagrammed in Figure 4. It can be seen that the defensive parts of the Bishop loci pass through the Cournot point CC; this is always the case, for $x_2(CC)/x_1(CC) = x_2(SS)/x_1(SS) \equiv r$. It follows that the points BC, CB, and CC are coincident for all a in the closed unit interval. In the diagram the points BS and SS are coincident; this happens only when $a \leq 1/2$. On the other hand, if $a \geq 1/2$, $SB = SS$.

Any combination of the above five strategies is also a strategy. The firm could, for example, follow a Stackelberg strategy for a while and, if that proved unprofitable, switch to the Bishop strategy. In fact, if the firm were really defensive-minded, it could switch from strategy to strategy in a random way, keeping its rival off balance. But such physical mixing of strategies is alien to the notion that firms *want* their outputs to be predictable, so long as their rivals use their predictions to behave in the proper way. Of course, a general theory of duopoly would yield this kind of behavior as a deduction from the conditions of the game, but I am not chasing that chimera.

III. Keeping Score

Straightforward but tedious computations yield the profits associated with the twenty-four labeled points in Figure 4. But these are not necessarily the profits associated with simultaneous pursuit of a given pair of strategies. Much depends on outputs at the beginning of the game, and the speed of their convergence to an intersection point, relative to the length of the accounting period. Thus, if both firms select their initial Bishop outputs near their monopoly points, the result in the first instance will be very much like that at SS, and the firms may take so long learning each other's intentions that the accumulation of profits over the learning period, plus the one-period profits at BB, fall short of those which could be earned during a convergence to the Cournot point. If we then conceptually "stop the game" and write down the score at that moment, we must write $P_i(CC) > P_i(BB)$, $i = i$, 2. But if we let the game continue for, say, twice as long, and then stop it, we might have to write $P_i(BB) > P_i(CC)$. If one or more accounting periods pass before an intersection is reached, how are the payoffs evaluated?

I have tried to handle this problem in the simplest possible way. No accounts are kept prior to the achievement of an intersection: those, and only those, profits earned at intersections are counted. If you like, we may imagine a Walrasian *tatonnement:*—firms come to market and announce their supplies, but no transactions take place until the same magnitudes have been announced a sufficient number of times. Or, in a less far-fetched vein, we may imagine the game to have continued for some time, the firms to have settled down to a definite strategy, and the returns associated with the movements *to* their present positions at an intersection to have been swamped by the returns being earned *at* the intersection. This means we are analyzing only the *equilibrium* properties of the strategies, but we

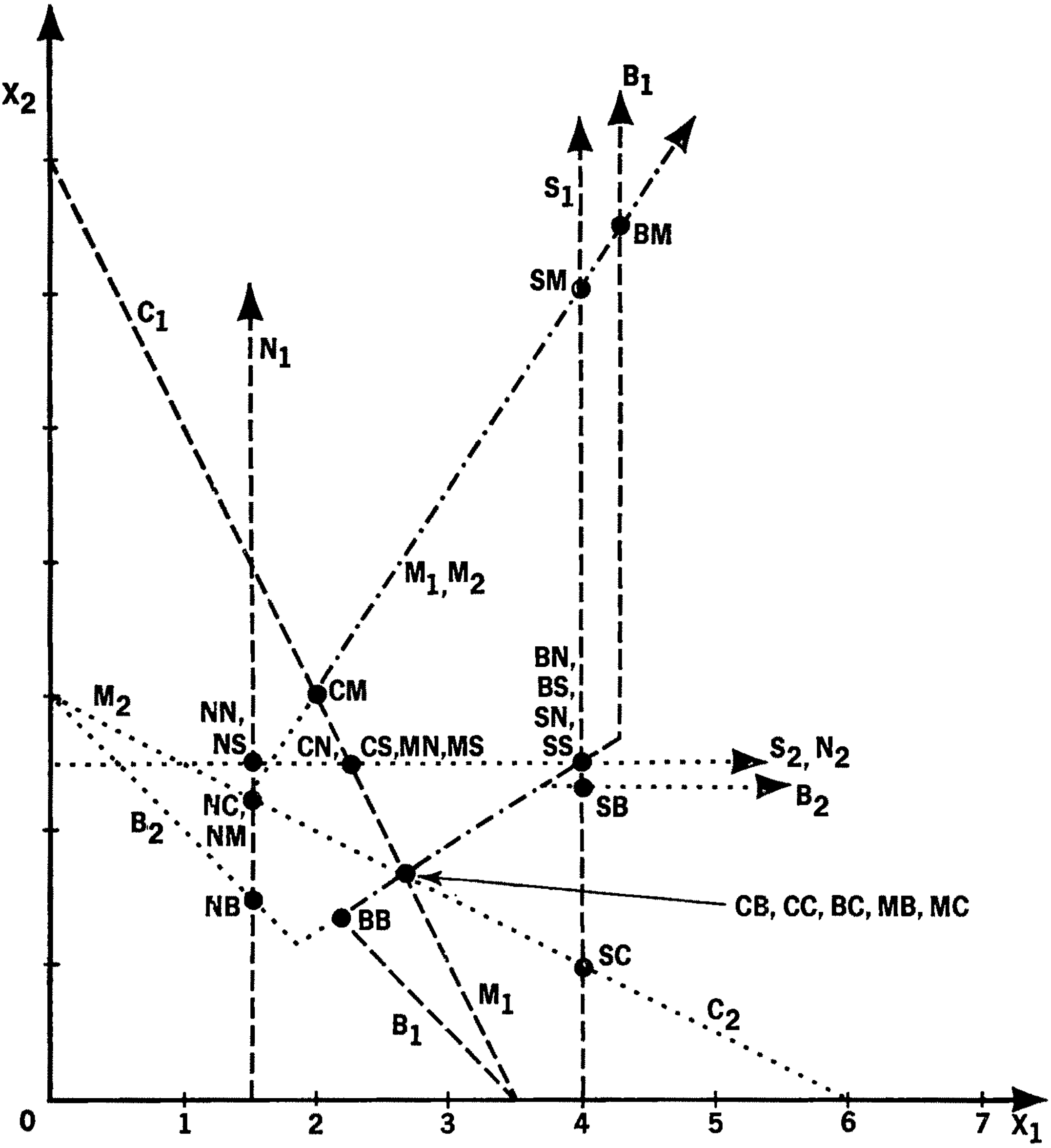


FIGURE 4. ($a=0, D=7, k=1.5$)

are already restricted in this regard by the proposed solution.[19]

It is necessary to be more explicit about this matter of the length of the game. The market conditions set forth in equations

---

[19] Whether equilibrium theory is the right way to study duopoly is a profound question. An equilibrium of the sort studied here may never be reached, in which case all the behavior escapes the analysis. But William Baumol, for example, has observed the slight attention paid by oligopolists to their rival's possible reactions. This would seem to indicate the common existence of strategic equilibria. Moreover, the modern theory of the firm as developed by Penrose, J. Downie, and Marris almost wholly ignores the question of oligopolistic reactions. This apparently rests on the implicit assumption that oligopolistic interdependence is a minor problem for the firm. I agree with this outlook so far as it indicates the easy attainability of price and output agreements, but even so, I remain very uncomfortable about the limitations of equilibrium.

(1)–(3) are supposed to have persisted for a long time. The firms have already gone through the trying business of feeling each other out (assessing will power, learning about likely reactions, etc.) and have reached some sort of mutually tolerable equilibrium. Possibly—indeed, probably—one of the firms would prefer an equilibrium more favorable to itself, but in view of past history it thinks a real improvement is unlikely. This state of affairs could go on indefinitely. Then something happens to change the symmetry relations. Perhaps one firm increases its efficiency relative to the other. This need not happen suddenly, but its perception—or at least the decision to act on it—is sudden. Now at least one of the firms finds it advisable to reopen the question of its market policy, and makes the other firm aware of this by any of a number of means. The firms must begin anew the difficult process of finding a mutually acceptable *entente*. The long submerged conflict reemerges. The game begins.

We, remaining resolutely statical in our outlook, do not appear on the scene until the new equilibrium is reached, at which time we record results and observe the character of the outcome. The game continues in the sense that moves continue to be made; but it has ended in the sense that equilibrium strategies have been found. The firms have "solved" their duopoly problem, and for us the game is over.

IV. The Set of Good Strategies

A strategy T_1 *dominates* a strategy T_2 if for any choice of strategy U_j by an opponent, profits associated with T_1 are not less than those associated with T_2. Clearly, no "best response" strategy is dominated; hence no Nash equilibrium point (hereafter, equilibrium point) occurs in dominated strategies. Thus good strategies are not dominated; but a strategy is not good simply because it is not dominated. Suppose a strategy is defined by a "conjectural variation" (a la Frisch); if this conjectural variation is certain to be false the strategy is not good.[20] We may define a strategy to be *good* if it is not dominated and is not based on certainly false conjectural variations.

The second condition defining good strategies lacks the intuitive force of the first, but in the interest of space I shall not justify it (beyond admitting a reluctance to imagine firms basing their plans on expectations they *know* will not be realized). The main reason it is advanced is that of expositional simplicity. It will develop that a firm's N (Nichol) strategy is not automatically dominated whenever its rival's M (Market Share) strategy is. If, consequently, the good strategies were regarded as merely the undominated ones, it would be necessary to determine the critical k (Output Ratio) at which the N strategy becomes dominated. This critical k is of no further use in the analysis because of the theory of solution adopted here (i.e., if M_j is dominated, N_i is not a member of an equilibrium point) and it doesn't seem particularly interesting on its own merits. Therefore the exposition is simplified at no great cost if it can be neglected. The suggested definition of good strategies permits this.[21]

Conditional Inferiority of the N Strategy. Since it is a best response to an M strategy, the N strategy will not be dominated when the former isn't. When the rival's M strategy *is* dominated, however, the conjectural variation defining the N strategy will certainly be false. Consequently the N_1 strategy is good only when the M_2 strategy is, and the N_2 strategy is good only when M_1 is.

---

[20] If a strategy of firm i is posited on the supposition that $dx_j/dx_i=v$; v is the conjectural variation. Possibly v may turn out to have been a false conjecture, but it is not certainly false *ex ante* unless it is certain that $dx_j/dx_i \neq v$. Of our strategies, C, N, and S are defined by conjectural variations—e.g., N_1 is posited on $dx_2/dx_1=k$ or on $dx_2/dx=dx_2(C)/dx_1$. Clearly if M_2 is dominated, N_1 is based on a certainly false conjectural variation.

[21] Alternatively, if one wanted to assume cardinal utility, and consequently to consider mixed strategies, one could show that when M_j is dominated by C_j, N_i is dominated by a mixture of B_i, C_i, and S_i (for the latter are all the best responses to the remaining j strategies).

Conditional Inferiority of the M Strategy. The M strategy is not defined by a conjectural variation; hence it will be good unless it is dominated. To show analytically when M is dominated requires the computation of forty-eight entries of the 5x5 payoff matrix (twenty-four for each firm, profits at MM being indeterminate). It will save a great deal of space, however, if the reader will accept the following assertion: to see when M is dominated it is only necessary to compare its profits with those of C. To be sure, other strategies will dominate M at some k, but only when C does. Hence we need only compare $P_1(CJ)$ with $P_1(MJ)$, and $P_2(JC)$ with $P_2(JM)$, for $J=B,C,M,N,S$. Furthermore, M_1 is dominated unless $P_1(MC)>P_1(CC)$. In place of an analytical proof of this proposition, the following verbal argument is offered.

> (i) $P_1(CB)\geq P_1(MB)$. If $k<r$, MB is found in the third phase of B_2, and that is not as good for firm one as CB; for CB lies in the defensive phase of B_2 and consequently on a lower (read: better) iso-profit contour.[22] On the other hand, if $k\geq r$, MB and CB are the same point.
>
> (ii) $P_1(CN)\geq P_1(MN)$ and $P_1(CS)\geq P_1(MS)$. Thus $x_2(S)$ is constant, and for given k, so is $x_2(N)$; and C_1 is defined as the locus of profit maximizing x_1's for constant x_2.
>
> (iii) $P_1(CM)$ vs. $P_1(MM)$. If $k\leq r$, CM is found at CC. MM could lie anywhere on the coincident parts of the M loci, but the further to the left it lies, the higher the profits of firm one. The extreme leftward point on the coincident parts of the M loci is MC (see Figure 1). Hence, to compare $P_1(CM)$ with $P_1(MM)$ when $k\leq r$, we may compare $P_1(CC)$ with $P_1(MC)$. This will give MM every benefit of the doubt. On the other hand, if $k>r$, CM is the extreme left-most point on the coincident parts of the M loci, so $P_1(CM)\geq P_1(MM)$. In summary, $P_1(CM)\geq P_1(MM)$ unless $P_1(MC)>P_1(CC)$.

These inequalities can be proved by solving simultaneously the appropriate equations defining the strategies, putting the solutions into equation (4), and comparing the results. Alternatively, Figures 1 and 5 can be studied.

From equations (A2) and (A8) of the Appendix, one sees that the MC and CC profits of firm one will be equated by

$$(24)\quad k_1 \equiv (D + 9a - 5)/4(D - 3a + 1)$$

and by $k\geq r$. Consequently C_1 dominates M_1 unless $k\in K_1^M$, where

$$(25)\qquad K_1^M \equiv \{k\colon k_1 < k < r\}$$

The interpretation of this result is straightforward. If the firm's market share is "very small" it is not profitably maintained; some other strategy is bound to yield greater profits and possibly a greater market share. Nor is the share profitably maintained if it is "very large"; a different strategy is bound to yield greater profits but probably a smaller market share. Some market shares are not irrationally maintained, in the sense that maintaining them is not a dominated strategy. The proper or minimally acceptable share ($k=r$) is not one of these—or at any rate not one reasonably maintained by the M strategy; it is better maintained by the B strategy.[23]

The meaning of very small and very large in this respect is provided by K_1^M. Too much should not be made of the pre-

---

[22] Firm one's iso-profit loci are shaped rather like inverted U's, with maxima lying on the C_1 locus, and represent greater profits the lower they are in output space. Firm two's are rather reverse C-shaped, and are "better" the more to the left they are.

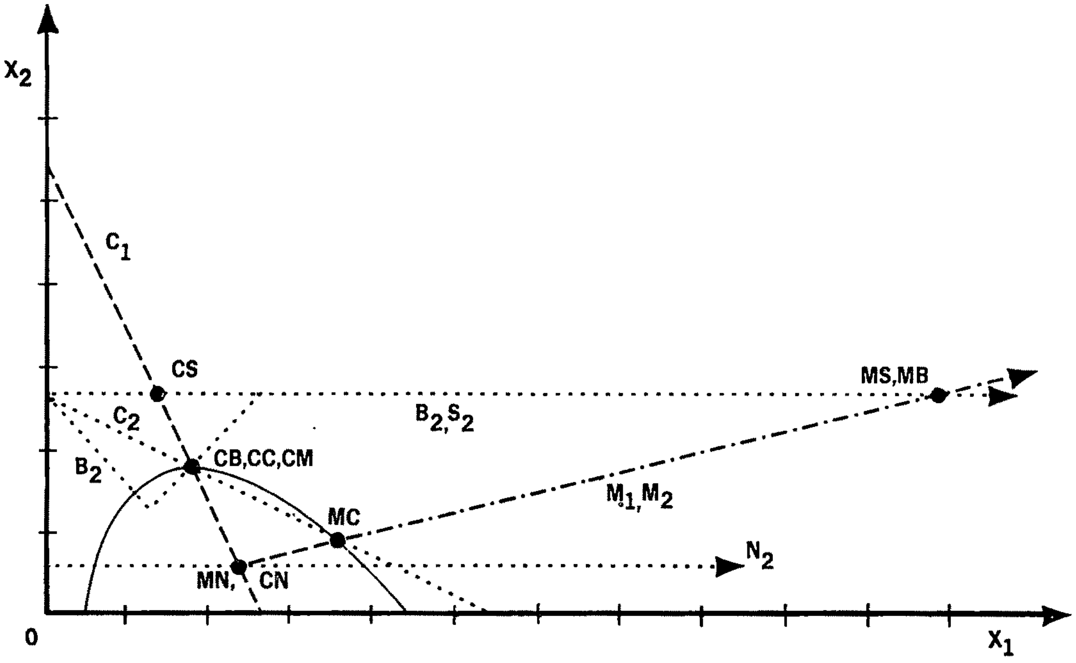


FIGURE 5. FIRM 2'S STRATEGIES AND FIRM 1'S C AND M STRATEGIES ($a=1/2$, $D=6$, $k=1/4$)

cise limits of this set because of the linearities assumed here. However, it is interesting to note that when the duopolists' costs are approximately equal, r is near unity, and hence the first firm would only play its M strategy when it is the larger firm (in terms of sales). For the smaller firm, M is a dominated strategy.[23]

The preceding analysis could now be repeated from the standpoint of the second firm, with symmetrical results. This firm's M strategy is dominated if $k<r$ and if $k \geq k_9$. The critical k_9 is found most easily by substituting $1-a$ for a in equation (24) and inverting. The result is

$$k_9 \equiv 4(D+3a-2)/(D-9a+4) \tag{26}$$

Letting

$$K_2^M \equiv \{k: r<k<k_9\}, \tag{27}$$

one sees that M_2 is dominated unless $k \in K_2^M$. Notice that the M strategies are not undominated at the same k, as K_1^M and K_2^M are disjoint.

When M is not dominated, does it dominate B,C, S, or N? No; it is a worse response than C to S, than S to C, than B to B, than N to M. Furthermore C_i is a best response to S_j, S_i is a best response to C_j, and B_i is a best response to B_j. Hence, B,C, S are never dominated, and consequently neither C nor S are defined by certainly false conjectural variations.

We have now shown that G, the set of good strategies, is

---

[23] As it is defined, k_1 could be negative. But that would be meaningless; both the numerator and denominator of k_1 (and all other k's as well) should be defined more precisely, as the maximum of zero and the quantity as given in the text, to make them nonnegative. Observe that unless k_1 as defined is negative, it would be irrational of the first firm to try to hold on to 100 percent of the market—to set a 'limit price.' Even then, though it would not certainly be irrational to try to exclude entry, it would not certainly be the best thing to do (M never dominates). (There is no requirement that both firms actually be in the market—one could be a potential entrant. Thus in analyzing the goodness of the M strategy we get as a by-product the beginnings of an analysis of the limit-pricing theory.)

$$(28)\quad G\begin{cases} = G^* + M_1 + N_2 & \text{if } k \in K_1^M, \\ = G^* + N_1 + M_2 & \text{if } k \in K_2^M, \\ = G^* \text{ otherwise,} \end{cases}$$

where

$$(29)\quad G^* = \{B_i, C_i, S_i, \text{ for } i = 1, 2\}$$

V. Market Properties of the Game-Theoretic Solutions

Equilibrium Points. From the definitions of the Bishop, Cournot, and Stackelberg strategies it follows that BB, CS, and SC are equilibrium points. The possible remaining equilibrium points are NM and MN. Clearly, the Nichol strategy is the best response to the Market Share Strategy. The converse is only true, however, when N intersects the corner of the M locus, in which case M and C are both best responses to N; in all other cases, C is the best response. Considering NM first, one sees from equation (15) that NM is a corner on M_2 when $q<k<q_1$ and $a<1/2$, and when $q<k$ and $a \geq 1/2$. But M_2 certainly won't be played unless $r<k<k_9$. Since $r>q$, NM is an equilibrium point if either:

$$(30)\quad r < k < \min[k_9, q_1] \quad \text{and} \quad a < 1/2,$$

$$(31)\quad r < k < k_9 \quad \text{and} \quad a \geq 1/2$$

Let (30) and (31) define K_{NM}; i.e.,

$$(32)\quad K_{NM} \equiv \{k: k \text{ satisfies (30) or (31)}\}$$

NM is an equilibrium point if $k \in K_{NM}$.

Turning to MN: M_1 will be good only when $k \in K_1^M$ and will be a best response to N_2 when intersected at its corner by N_2. According to equation (16) this will happen when $s_1 \leq k \leq s$ and $a>1/2$, and when $k<s$ and $a \leq 1/2$. Since $r<s$, MN is an equilibrium point when, on the one hand, $k \in K_1^M$ and $a \leq 1/2$, and on the other, when

$$(33)\quad \max[k_1, s_1] < k < r \quad \text{and} \quad a > 1/2$$

Hence, MN is an equilibrium point when $k \in K_{MN}$, where

$$(34)\quad K_{MN} \equiv \{k: k \in K_1^M \text{ and } a \leq 1/2, \text{ or } k \text{ satisfies (33)}\}$$

Hence the set of equilibrium points, $\mathcal{E}$, is

$$(35)\quad \mathcal{E} = \{BB, CS, SC \text{ and: } NM \text{ if } k \in K_{NM}, \quad MN \text{ if } k \in K_{MN}\}$$

This, then, answers one of the questions posed near the beginning of the paper. Every equilibrium point of the present duopoly output game is a solution proposed by a duopoly theorist; and conversely, the Bishop, Stackelberg, and Nichol theories represent equilibrium points. We observe in passing that the set of equilibrium points covers those outcomes where conjectural variations are accurate.

Maximin. A study of the diagrams will help the reader to see, what I don't want to take the space to prove, that the Cournot strategy is a "safe" one. It is intuitively obvious that a plan of always choosing the profit-maximizing output, given the rival's output, cannot lead to the *greatest* difficulties. The worst that can happen from following this strategy is better than the worst outcomes to the B, M, or S strategies, and is usually better than the worst outcome to the N strategy. The minimal outcome to playing C_1 is found either at CS or CM, depending on k, and this exceeds the minimal outcome to the N_1 strategy (found at NS) unless k is near the upper limit of K_2^M. In the latter event the worst Cournot result is at CM, which is *worse* than the worst Nichol result at NS. Hence in order to determine when N_1 is maximin it is necessary to compare $P_1(CM)$ with $P_1(NS)$ for $k \in K_2^M$. When the latter profit is larger, N_1 is maximin. The critical k at which this happens can be found by using equations (A9) and (A10). Without computing it, call this critical value k_8; i.e.,

$$(36)\quad k_8 \equiv k \text{ such that } P_1(CM) = P_1(NS) \quad \text{for } k \in K_2^M$$

It can be shown that when $k_8 \leq k < k_9$, N_1 is maximin. Of course for these k, M_1 is dominated and hence N_2 is not good; therefore C_2 is the second firm's maximin strategy. Consequently the maximin point is NC.

Applying symmetry and defining

$$k_2 \equiv k \text{ such that } P_2(MC) = P_2(SN) \quad \text{for } k \in \overset{M}{K_1}, \tag{37}$$

it can be shown that CN is the maximin point when $k_1 < k \leq k_2$. In all other cases CC is the maximin point. In summary, the set of maximin points, $\mathfrak{M}$, is

$$\mathfrak{M} \begin{cases} = CN & \text{if } k_1 < k \leq k_2, \\ = NC & \text{if } k_8 \leq k < k_9, \\ = CC \text{ otherwise}^{24} \end{cases} \tag{38}$$

The critical k_2 and k_8 are of no further use in the analysis and so will not be derived. In any event the most interesting aspect of this result is the fact that, if both firms play C they will end up at CC and, consequently, k will equal r—both M strategies will be dominated and so *neither* N strategy will be good. In playing C, then, both firms will be playing their maximin strategies.[25] It is not that the strategies are played because they are maximin; rather, they are maximin because they are played. This illustrates a curious feature of duopoly games: the character of a strategy is in some measure determined by whether it is being played, and if not being played, by what strategies *are* being played. We may observe, additionally, that when D is very large the firms' costs will be negligibly small in relation to market demand, and we will have Cournot's mineral spring problem. But for that situation it can be shown that CC is certainly the maximin point. The oldest and the newest duopoly solutions are the same!

Joint Maximum. Because it is much closer than the other points to the contract curve, BB is the joint maximum point in G^*. It is usually the same in G too, but not always. There are k for which NB (or BN, as the case may be) is closer to the contract curve. For this it is necessary that $a \neq 1/2$, for otherwise BB is *on* the contract curve. Even so, the range of k for which BB is not the joint maximum is very small. I shall not bother to compute the range, for the joint maximum is without significance in the present analysis. Indeed, a joint maximum of ordinal utilities exists only when there is a point at which both firms' utilities are greatest. Even this is not sufficient unless interpersonally comparable utilities are assumed. And with all that no behavioral significance is gained unless side payments are permitted. Such information as is given here about the joint maximum is offered purely as a curiosum.

Economic Efficiency of the Game-Theoretic Solutions. Of the points identified in equations (35) and (38), BB, being the most efficient for the duopolists, is in most cases the least efficient for consumers. All the points are inefficient (price exceeds the lowest marginal cost) but their ranking in this regard depends on the state of cost inequality and market shares (and on the

---

[24] These are not saddle points; such exist only in mixed strategies.

[25] For G^*, the payoff matrices are satisfied by the following, where the first numeral of each pair in parentheses is the first firm's profit.

for $a<\frac{1}{2}$	B_2	C_2	S_2
B_1	(5, 4)	(4, 2)	(1, −1)
C_1	(4, 2)	(4, 2)	(3, 3)
S_1	(2, 0)	(6, 1)	(1, −1)

for $a=\frac{1}{2}$	B_2	C_2	S_2
B_1	(3, 3)	(2, 2)	(0, 0)
C_1	(2, 2)	(2, 2)	(1, 3)
S_1	(0, 0)	(3, 1)	(0, 0)

The order in firm one's (and respectively, firm two's) profits shown by this example are completely independent of a and D (if all outputs are positive) and largely independent of the linearities in costs and demand.

definition of inefficiency). Consider first the case of equal costs. Then the efficiency locus, i.e., $\{x_1, x_2: p(x_1, x_2)=1/2\}$, joins the ends of the C loci which are opposite to the respective monopoly points. It requires no formal argument to see that SC and CS are the least *inefficient* points (price is closest to marginal cost), CC the next least inefficient, followed by NM and NC, and finally BB. Both most and least inefficient points are equilibrium points; the maximin point is intermediate in inefficiency.

If costs are not equal the efficiency locus reduces to the point on the lowest-cost firm's output axis where this firm's marginal cost equals price (see Figure 4, where the efficient point is $x_1=7$, $x_1=0$). One measure of inefficiency in this case is the distance of a point from the efficient point. According to this measure the least inefficient point is the low-cost firm's Stackelberg leadership point SC, followed by CC, then by BB, MN, and NM. But another measure of inefficiency can be imagined—namely, inefficiency increases as the total output decreases. According to this measure the least inefficient point is again the low-cost firm's Stackelberg point, followed by the high-cost firm's Stackelberg point, followed by CC, then by a Nichol point; most inefficient is the Bishop point. Hence on either measure an equilibrium point is least inefficient and another equilibrium point is least *efficient*, and the maximin point is intermediate in inefficiency.

In other words, *the relative economic inefficiencies of game solutions are not indicated by the solution concept adopted.*

VI. The Proposed Solution and Concluding Comments

It would overburden an already long exposition to display the complete set of points that are Pareto optimal to the firms, so I shall simply derive the Pareto optimal equilibrium points—that is, the undominated equilibrium points. For this it is necessary to compute and compare profits at the Nichol, Bishop, and Stackelberg points. Naturally, those comparisons involving Nichol points depend on k. That is, for some k, $P_1(MN)>P_1(SC)$, say, but for some other k, $P_1(SC)>P_1(MN)$. However, in comparing MN with SC we must put $k=q$, the market ratio at SC. That is to say, if at some point in the course of play $k\neq q$ and both firms view SC as being superior to MN (based on the current k) and if they select the S and C strategies, respectively, they will arrive at SC, thus causing k to become equal to q. But the firms will not remain at SC if a Nichol point gives both of them greater profits, for on our assumptions the game does not end—the market does not settle down to equilibrium—until all possibilities of mutual improvement have been exhausted. Therefore, having arrived at SC, and thereby having changed k, the firms will reopen the question of their optimal strategies. (In point of fact, of course, this question must be kept open as long as k is changing.)

Now it can be shown that when $k=q$, both firms obtain greater profits at MN than at SC: MN Pareto-dominates SC. On the present theory of solution, then, SC is not a solution; and by symmetry, neither is CS—it is Pareto-dominated by NM. Note that this hypothetical "movement" from SC to MN (or from CS to NM) requires a reversal in the leader-follower relationship: in moving from SC to MN, the first firm abandons its role as Stackelberg leader in favor of that of a Nichol follower (maintaining market share). If a firm prized leadership this reversal of roles could block the movement. However, on our assumptions a firm seeks only profits, not leadership. Hence, neither Stackelberg point is a solution.

Now we must compare profits at the

$$(39)\quad \begin{cases} k_5 \equiv 2\left[\sqrt{\dfrac{2D-1}{D-3a+1}}-1\right], \\ k_6 \equiv \dfrac{(D+3a-2)\sqrt{D-a}}{2[(D+a-1)\sqrt{2D-1}-(D+3a-2)\sqrt{D-a}} \end{cases}$$

remaining equilibrium points: BB and either MN or NM. However, we do *not* put k equal to its value at BB—i.e., we do not put $k=r$. When $k=r$, BB Pareto-dominates both NM and MN, simply because both M strategies are dominated; but the reason we put $k=q$ in comparing MN and SC was the Pareto-dominance of MN over SC—the reverse, so to speak, of the present case. In other words, if the firms should arrive at BB, they would remain there, for BB Pareto-dominates NM and MN; but if they should arrive at NM (or MN) when $k\neq r$, they might well remain *there*, for BB does not Pareto-dominate NM and MN for all k. Or to put this in a different way: having arrived at SC, the firms will necessarily find a further movement (to MN) jointly beneficial; and having arrived at BB, they will certainly *not* find a further movement jointly beneficial; but having arrived at NM or MN they may or may not find a further movement jointly beneficial, depending on k. Hence MN (or NM) may be a solution.

If we define k_5 and k_6 as in equation (39), and refer to equations (21), (24), and (26), defining r, k_1, and k_9, we could show that the following hold:

$$\begin{array}{lllll} (40) & \text{in } P_1: & MN > BB & \text{if } k_1 < k < k_5, \\ & & BB > MN & \text{if } k_5 < k < r, \\ & & BB > NM & \text{if } r < k < k_9; \\ (41) & \text{in } P_2: & BB > MN & \text{if } k_1 < k < r, \\ & & BB > NM & \text{if } r < k < k_6, \\ & & NM > BB & \text{if } k_6 < k < k_9 \end{array}$$

(In (40) and (41) we use the fact that M_1 is undominated only if $k_1<k<r$, and M_2 is undominated only if $r<k<k_9$.)[26]

From these relations it follows that the set of Pareto optimal equilibrium points—the solution, $\mathcal{S}$,—is

$$(42)\quad \begin{array}{ll} \mathcal{S} = \{BB \text{ and: } MN & \text{if } k_1 < k < k_5, \\ \qquad\qquad NM & \text{if } k_6 < k < k_9\} \end{array}$$

It may be observed that the solution does not depend on a; whatever the relations between the firms' marginal costs, (42) is the solution. This would not be true if the M strategy were not considered. In that case BB would be the unique solution if $a=\frac{1}{2}$, and would be joined by either SC or CS if $a\neq\frac{1}{2}$.[27]

Implications for Economic Policy. For all a in the closed unit interval and all D large enough that MN and NM outputs are positive, $0<k_1<k_5<r<k_6<k_9$. Hence the solution will be unique only when k is *not* in either open interval (k_1, k_5), (k_6, k_9). We could find a measure of the likelihood of k being in these intervals, and ask after social policies which would increase it, or at least not decrease it. These would be good rules, in that they would increase the likelihood of the less inefficient Nichol point being in the solution.

But there is a problem here. The k that we have been discussing is the one pertinent to the course of play; but the k pertinent to social policy, if duopolists usually reach equilibrium quickly, is the one that

---

[26] For the reader beginning to get lost in the forest of k's, I note that, generally speaking $k_i<k_j$ if $i<j$, but this order is not invariant with respect to a.

[27] See the matrices of fn. 25.

emerges from the play—the solution k. Thus, if the firms adopt a Bishop solution, k will equal $r \equiv (D+3a-2)/(D-3a+1)$. Conversely, if $k=r$, and k is stable over time, the firms have reached a Bishop solution. If two actual duopolists have approximately equal average variable costs ($a \cong \frac{1}{2}$) then $r \cong 1$, and approximate equality in their sales ($k \cong 1$) is evidence of the Bishop solution; but a significant inequality in their sales is evidence of the Nichol solution. Reasonably competitive factor markets would keep a from being very far from one half, and so generally speaking a k of near unity probably indicates the Bishop solution. But if for some reason the duopolists' costs were very different a k of near unity would indicate the Nichol solution; only very unequal sales (with the more efficient firm selling more) would indicate the Bishop solution.

We have to ask, then, what policies will be likely to prevent $k=r$. Actually, it is probably asking too much to expect policies to change a Bishop solution to the Nichol; the best to be hoped for is an avoidance of policies which would encourage the firms to adopt the Bishop solution. That is, a good practical policy would aim at keeping unequal k and r unequal. In terms of costs and market shares, approximately equal variable costs should be accompanied by unequal market shares, and rather unequal variable costs by rather more equal market shares. It would certainly be poor policy to favor the smaller of equally efficient firms, or the less efficient of equally large firms. Such measures, unfortunately, are typical of much legislation and judicial decisions in the area. On the present analysis they are mistakenly conceived.

This conclusion is reinforced by the speed with which equilibrium is likely to be reached when the costs and shares are equal. Not only is the Bishop point the unique solution in that situation, but it becomes a more obvious focal point as the situation is approached.[28] Periods of disequilibrium, during which the firms are finding their best strategies, and during which their outputs more closely approach the efficiency condition, are likely to be shortened as the firms more quickly recognize the obvious attractions of the Bishop point.

On the present analysis, with linear demand and cost functions, the Bishop outcome in conditions of cost equality is no better than the monopoly outcome: each firm produces exactly half what it would produce as a monopolist. Under more general conditions, which yield non-linear contract curves, the result in Bishop equilibrium would be slightly better than that in monopoly, but it may be doubted that firms look on their cost and demand functions as being very far from linear. There would seem little error in the conclusion that a duopoly in Bishop equilibrium offers no welfare gain over monopoly.

It is the worst sort of public policy, then, to seek to maintain equality in the market positions of quantity-adjusting duopolists, for it merely serves to foster the conciliatory behavior which produces the Bishop solution. *But it must be emphasized that this has not been established for the case of price-adjusting duopoly.*[29]

These remarks apply to resource allocation in static terms. Whether consumer welfare in general is better served at the Bishop point must also be considered. Might the firms more energetically seek product improvements and other beneficial forms of non-price competition at the Bishop than at the Nichol point? There is no clear evidence on this question, but it may be doubted that such is the case. One would expect the firm following the Nichol strategy, which, after all, is supplying less than its "proper" share of the market, to find greater incentive to better its position by the various methods of non-price competition. The most obvious and most commonly used method of achieving this is a real or supposed differentiation of products. This would seem a more likely result the greater the discrepancy between the firm's actual and proper market shares.

---

[28] Thomas Schelling presents evidence from informal experiments which seem to indicate the importance of focal points. See especially pp. 60–66.

[29] Nor has it been solidly established in precisely this form for output games. If the Market-share strategy had been defined simply as the maintenance of current market share without qualification, i.e., $x_1(M)=x_2/k$, $x_2(M)=kx_1$, then the Nichol strategy would become $x_i(N)=x_i^g$, and it could be shown that when C_i did not dominate M_i, S_i would (for D large enough to ensure positive x's). Neither M_i nor, therefore, N_i, would be good strategies for any k: the game would be reduced to G^*. The solution would be $\{BB$ and SC or $CS\}$ if $a \neq \frac{1}{2}$, and $\{BB\}$ if $a=\frac{1}{2}$. Since price is higher at BB than at the other solution points, good policy would attempt to keep $a \neq \frac{1}{2}$—to keep the duopolists variable costs unequal. The relation between a and k would not matter.

Successful differentiation of a firm's product introduces a new element into the duopoly game—a possible asymmetry in demand. The new situation is most naturally analyzed with price as the choice variable. That is the subject of a future paper.

Appendix

In this appendix are found those payoffs which are explicitly referred to in the text. Space limitations prohibit the derivation of all the entries in the 5x5 payoff matrix.

From equations (22) and (23) or from inspection of Figure 3, it can be seen that the Bishop point BB is the intersection of the low-cost firm's conciliation locus and the high-cost firm's defensive line (the line $k=r$). For $a \leq 1/2$, then, BB is yielded by the simultaneous solution of the first form of (22) and the second form of (23). But if $a \geq 1/2$, BB is found by the simultaneous solution of the second form of (22) and the first form of (23). In either case, the Bishop point outputs determine the Bishop point profits by equations (4) and (5). Hence, equation (A1) shows the values of P_1 and P_2 at BB. Proceeding in the obvious way, we find the values of P_1 and P_2 at the other solution points, as shown in A(2) through A(10).

Equations (A1) through (A7) may be used

$$\text{(A1)}\quad \begin{aligned} &\text{at } BB\text{: if } a \leq 1/2, P_1 = \frac{(D-a)^2(D-3a+1)}{4(2D-1)}, \qquad P_2 = \frac{(D-a)(D+3a-2)^2}{4(2D-1)}; \\ &\qquad\text{if } a \geq 1/2, P_1 = \frac{(D+a-1)(D-3a+1)^2}{4(2D-1)}, P_2 = \frac{(D+a-1)^2(D+3a-2)}{4(2D-1)} \end{aligned}$$

(A2) at $BC=CB=CC$: $P_1=(D-3a+1)^2/9$, $P_2=(D+3a-2)^2/9$;

(A3) at CS: $P_1=(D-5a+2)^2/16$, $P_2=(D+3a-2)^2/8$;

(A4) at SC: $P_1=(D-3a+1)^2/8$, $P_2=(D+5a-3)^2/16$;

(A5) at SS: $P_1=(1-2a)(D-3a+1)/4$, $P_2=(2a-1)(D+3a-2)/4$;

(A6) at BS: if $a \leq 1/2$, $BS=SS$;

$$\text{if } a>1/2, \quad P_1=\frac{(1-2a)(D-a)(D-3a+1)(D+3a-2)}{2(2D-1)^2}$$

$$P_2=\frac{3(D-a)(D+3a-2)}{4(2D-1)};$$

$$\text{(A7) at } SB\text{: if } a<1/2, \quad P_1=3(1-2a)(D+a-1)(D-3a+1)/4(2D-1),$$
$$P_2=(2a-1)(D+a-1)(D-3a+1)(D+3a-2)/2(2D-1)^2$$
$$\text{if } a\geq 1/2, \quad SB=SS;$$

$$\text{(A8) at } MC\text{: if } k<r, \quad P_1=\frac{(D+a-1)[1-2a+k(D-3a+1)]}{(2k+1)^2},$$
$$P_2=\left[\frac{k(D+a-1)}{2k+1}\right]^2,$$
$$\text{if } k\geq r, \quad MC=CC;$$

$$\text{(A9) at } CM\text{: if } k\leq r, \quad CM=CC,$$
$$\text{if } k>r, \quad P_1=[(D-a)/(k+2)]^2,$$
$$P_2=\frac{k(D-a)[D+3a-2+k(2a-1)]}{(k+2)^2};$$

$$\text{(A10) at } NS\text{: if } k\leq q, \quad NS=SS,$$
$$\left.\begin{array}{l}\text{if } q<k\leq q_1 \text{ and } a<1/2,\\ \text{or if } q<k \text{ and } a\geq 1/2,\end{array}\right\} P_1=\frac{(D+a-1)[(2k+1)(D-5a+2)-2(D+a-1)]}{2(2k+1)^2};$$
$$\text{if } q_1<k \text{ and } a<1/2, \quad P_1=\frac{(D-a)[k(D-5a+2)-(D+4a-2)]}{4(k+1)^2}$$

---

to prove that the orders in P_1 and P_2, respectively, shown in footnote 25, are indeed invariant for all a in the closed unit interval and all D large enough that the outputs in question are positive. Thus, equation (7) indicates that $D-3a+1$ must be positive for $x_1(S)$ to be positive, so in proving that $P_1(BC)>P_1(SS)$ we divide the two profits by $D-3a+1>0$ (which will certainly hold if $D>2$) and find that the stated order in the two profits is implied by $4D+6a-5>0$—a relation true for all $D>2$ and all a in [0, 1]. The relations *between* P_1 and P_2, on the other hand, are without significance unless interpersonal comparisons of utility are made.

REFERENCES

W. J. Baumol, *Business Behavior, Value and Growth,* New York 1959.

R. Bishop, "Duopoly: Collusion or Warfare?" *Amer. Econ. Rev.,* Dec. 1960, *50,* 933–61.

R. Dorfman, P. Samuelson, and R. Solow, *Linear Programming and Economic Analysis,* New York 1958.

J. Downie, *The Competitive Process,* London 1958.

D. Ellsberg, "Theory of the Reluctant Duelist," *Amer. Econ. Rev.,* Dec. 1956, *46,* 909–23.

W. Fellner, *Competition Among the Few,* New York 1949.

F. M. Fisher, book review of *Business Behavior, Value and Growth,* by Baumol; *J. Polit. Econ.,* June 1960, *68,* 314–15.

R. Frisch, "Monopole, Polypole—The Concept of Force in the Economy," *International Economic Papers,* I, New York 1951.

J. Harsanyi, "A General Solution for Finite Non-Cooperative Games, Based on Risk-Dominance," in M. Dresker et al., eds. *Advances in Game Theory,* Princeton 1964, pp. 651–79.

A. Henderson, "The Theory of Duopoly," *Quart. J. Econ.,* Nov 1954, *68,* 565–84.

J. Kemeny and G. Thompson, "The Effect of Psychological Attitudes on the Outcomes of Games," *Contributions to the Theory of Games,* III, Princeton 1957, pp. 273–98.

A. Koo, "Recurrent Objections to the Minimax Strategy," *Rev. Econ. Statist.*, Feb. 1959, *41*, 36–41.

R. Luce and H. Raiffa, *Games and Decisions*, New York 1957.

R. Marris, *The Economic Theory of 'Managerial' Capitalism*, London 1966.

J. Nash, "Non-cooperative Games," *Annals of Mathematics*, Sept. 1951, 54, 286–95.

H. Neisser, "Oligopoly as a Non-Zero Sum Game," *Rev. Econ. Stud.*, Oct. 1957, *25*, 1–20.

A. J. Nichol, "Professor Chamberlin's Theory of Limited Competition," *Quart. J. Econ.*, Feb. 1934, *48*, 317–37.

E. Penrose, *The Theory of the Growth of the Firm*, New York 1959.

T. Schelling, *The Strategy of Conflict*, Cambridge 1960.

M. Shubik, *Strategy and Market Structure: Competition, Oligopoly, and the Theory of Games*, New York 1959.

G. J. Stigler, "A Theory of Oligopoly," *J. Polit. Econ.*, Feb. 1964, *72*, 44–61.

J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, 3d ed., Princeton 1953.

The Private and Social Value of Information and the Reward to Inventive Activity

By JACK HIRSHLEIFER*

A number of recent papers[1] have dealt with the economics of information in a context in which each individual is fully certain about his own endowment and productive opportunities. In those papers, the individual is imperfectly informed only about his market opportunities, i.e., about the supply-demand offers of *other* individuals. In consequence, costly patterns of search for trading partners replace the traditional assumption of costless exchange.

This paper deals with an entirely different aspect of the economics of information. We here revert to the textbook assumption that markets are perfect and costless. The individual is always fully acquainted with the supply-demand offers of all potential traders, and an equilibrium integrating all individuals' supply-demand offers is attained instantaneously. Individuals are unsure only about the size of their *own* commodity endowments and/or about the returns attainable from their *own* productive investments. They are subject to technological uncertainty rather than market uncertainty.[2]

Technological uncertainty brings immediately to mind the economics of research and invention. The traditional position has been that the excess of the social over the private value of new technological knowledge leads to underinvestment in inventive activity. The main reason is that information, viewed as a product, is only imperfectly appropriable by its discoverer.[3] But this paper will show that there is a hitherto unrecognized force operating in the opposite direction. What has been scarcely appreciated in the literature, if recognized at all, is the *distributive* aspect of access to superior information. It will be seen below how this advantage provides a motivation for the private acquisition and dissemination of technological information that is quite apart from—and may even exist in the absence of—any social usefulness of that information.[4]

I. Foreknowledge Versus Discovery

Within the category of technological (as opposed to market) information, dif-

---

* Professor of economics, University of California, Los Angeles. This paper is an abbreviation of a report with the same title prepared for Western Management Science Institute, UCLA (1970b). The research at WMSI was supported by the National Science Foundation and the Office of Naval Research. Valuable suggestions have been contributed by Earl Thompson, Jacob Marschak, Ross M. Starr, Benjamin Klein, and Joseph Ostroy.

[1] See S. A. Ozga, George Stigler (1961, 1962), and Armen Alchian.

[2] These two types of uncertainty have been distinguished by a number of authors: see Tjalling Koopmans (pp. 161ff), Peter Diamond, Jacob Marschak (1968b, p. 17).

[3] See Kenneth Arrow (1962, p. 619). The comment by Harold Demsetz is also of interest. On patents as a device to achieve appropriability, see Fritz Machlup.

[4] In connection with policy debates over stock market "insider trading," Henry Manne has discussed the private and social gains attached to the dissemination of corporate information. A very recent article by Eugene F. Fama and Arthur B. Laffer emphasizes the differing motivations of insiders and outsiders for the generation of such information; their analysis, though in a partial-equilibrium context, in some ways parallels the treatment here.

ferent sorts of knowledge are associated with rather different private incentives and social efficiency conditions. This paper concentrates upon a distinction between what will be called foreknowledge on the one hand, and discovery on the other. The type of information represented by foreknowledge is exemplified by ability to successfully predict tomorrow's (or next year's) weather. Here we have a stochastic situation: with particular probabilities the future weather might be hot or cold, rainy or dry, etc. But whatever does actually occur will, in due time, be evident to all; the only aspect of information that may be of advantage is prior knowledge as to what will happen. Discovery, in contrast, is correct recognition of something that possibly already exists, though hidden from view. Examples include the determination of the properties of materials, of physical laws, even of mathematical attributes (e.g., the millionth digit in the decimal expansion of "π"). The essential point is that in such cases Nature will not autonomously reveal the information; only human action can extract it. Foreknowledge information is conceptually simpler to deal with, involving as it does only the value of *priority* in time of superior knowledge; this topic will be taken up first below.

II. Elements of the Economics of Foreknowledge

The analysis of the value of priority of information necessarily involves both temporality and uncertainty. For convenience, the simplest possible paradigm of choice will be employed. Suppose that there exists but a single physical good (corn). It will be assumed, however, that a number of different types of claims may be owned or traded—claims to corn at specified dates and under specified contingencies or "states of Nature."[5] It suffices to consider a particularly simple model in which the present (time 0) is certain, and the future consists of a single date (time 1) at which just one of two alternative states (a or b) will obtain. The marketable commodities of the analysis can be denoted c_0, c_{1a}, and c_{1b}: claims to corn valid at, and only at, the subscripted dates and states.

Each individual will have a utility function governing his preferences *now* for holdings of alternative combinations of these claims. Entering into this function will be his beliefs as to the probabilities π_a and π_b of the two states. It has been shown[6] that, under certain widely accepted assumptions, it is possible to find a cardinal intertemporal function $u(c_0, c_1)$ that (a) measures desirability of alternative *certain* dated consumption sequences and (b) is such that the von Neumann-Morgenstern expected-utility rule can be employed to order preferences among *risky* sequences of time-state claims, according to the formula:

$$(1) \qquad U = \pi_a u(c_0, c_{1a}) + \pi_b u(c_0, c_{1b})$$

This utility function exemplifies the property of "state-independence," i.e., the expected utility is a sum of distinct terms each of which is associated with only one particular state of the world.[7]

The utility function to be used below

---

[5] The conception of state-claims as commodities stems from the pioneering work of Arrow (1953, 1964). Gerard Debreu (ch. 7) extended Arrow's model to multiple time-periods. The paradigm of choice involving time-state claims has been further developed by other authors, including Karl Borch and Hirshleifer (1965, 1966).

[6] See Jacques Drèze and Franco Modigliani; Hirshleifer (1970a, ch. 8).

[7] *State-independence* is an implication of the von Neumann-Morgenstern postulate sometimes called "irrelevance of nonaffected outcomes"—see Marschak (1968a). The key idea is that when we are dealing with *prospects* which promise to offer one consequence if state a obtains and another if state b obtains, we need not consider any relations of complementarity in preference. For there is never any question of receiving the *combined* consequences attached to the two states; the individual will necessarily receive one to the exclusion of the other.

makes the further specification that $u(c_0, c_1)$ can be written in the special form $v(c_0)+\theta v(c_1)$, where θ is a fixed time-preference parameter characteristic of the individual, and v is a cardinal preference-scaling function for income valid for him at each state and date. The essential additional property underlying this specification may be called time-independence.[8] Given both state-independence and time-independence, the utility function becomes:

$$(2) \qquad \begin{aligned} U &= \pi_a(v_0 + \theta v_{1a}) + \pi_b(v_0 + \theta v_{1b}) \\ &= v_0 + \theta(\pi_a v_{1a} + \pi_b v_{1b}) \end{aligned}$$

where v_0 is condensed notation for $v(c_0)$, and v_{1a} and v_{1b} are defined analogously.

The acquisition of information will take the form of warranted revisions in the probability estimates π_a and π_b that enter into preference functions and so guide decisions. In what follows it will be essential to distinguish *private* information (available only to a single individual) from *public* information (available to everyone)—intermediate cases will generally be ignored. It will also be important to keep in mind the distinction between information that is *prior* to, and information that is *posterior* to, the individual consumption-investment decisions that must be made at t=0. Still another distinction is that between *sure* information (as to which future state will obtain) and merely *better* information—the latter would represent a sharper focusing of subjective probabilities that does not entirely eliminate uncertainty.

III. The Value of Foreknowledge: Pure Exchange

In a simplified world of pure exchange, all productive transformations among the quantities c_0, c_{1a}, and c_{1b} are ruled out—even simple storage. An individual dissatisfied with his endowment vector $Y = (y_0, y_{1a}, y_{1b})$ can modify it only by trading. Let us suppose a competitive world of "representative" individuals, characterized by identical probability beliefs and utility functions, and all holding identical endowments. Then no trading actually takes place; the price structure that emerges in market equilibrium must "sustain" the endowment pattern for every individual. Taking current corn as numeraire so that $P_0=1$, the sustaining prices must be:[9]

$$(3) \qquad P_{1a} = \theta\pi_a v_{1a}'/v_0' \quad \text{and} \quad P_{1b} = \theta\pi_b v_{1b}'/v_0'$$

Wealth in c_0 units may then be determined from the definition:

$$(4) \qquad W_0 = P_0c_0 + P_{1a}c_{1a} + P_{1b}c_{1b}$$

Finally, utility attained may be calculated by appropriate substitutions in (2). This is the base situation with which the results of changes in information will be compared.

For concreteness, a numerical illustration will be employed using a logarithmic preference-scaling function v (see Table 1). Note that future state a is assumed better endowed than state b ($y_{1a}=200$, $y_{1b}=80$) as well as more probable ($\pi_a=.6$, $\pi_b=.4$). With c_0 as numeraire so that $P_0=1$, the parameters assumed lead to the solutions shown in column (4) for the sustaining prices ($P_{1a}=.3$ and $P_{1b}=.5$) and wealth ($W_0=200$). The expected utility (U

---

[8] *Time-independence,* the absence of complementarity in preference between income at time 0 and income at time 1, does not have so compelling a justification as does state-independence. For the individual will indeed be receiving a combination of consequences over time. But, in the absence of any convincing reason for anticipating positive complementarity or its reverse, the assumption of zero complementarity may be a satisfactory simplification. The assumption is widely employed in the literature of intertemporal optimization (see, e.g., Arrow (1966, p. 20)).

[9] The individual maximizes $U=v_0+\theta(\pi_a v_{1a}+\pi_b v_{1b})$ subject to $P_0c_0+P_{1a}c_{1a}+P_{1b}c_{1b}=P_0y_0+P_{1a}y_{1a}+P_{1b}y_{1b}$. The usual Lagrangean conditions lead to $v_0'=\lambda P_0$; $v_{1a}'=\lambda P_{1a}/\theta\pi_a$: $v_{1b}'=\lambda P_{1b}/\theta\pi_b$. With $P_0=1$, and (since all individuals have identical preferences and opportunities) $c_0=y_0$, $c_{1a}=y_{1a}$, and $c_{1b}=y_{1b}$, the results in the text are obtained.

TABLE 1—DATA AND SOLUTION FOR NUMERICAL EXAMPLE

Endowments (1)	Prior Beliefs (2)	Preference Parameters (3)	Solution, Base Case (4)
$y_0=100$	—	$\theta=1$	$P_0=1$
$y_{1a}=200$	$\pi_a=.6$	$v=log_e c$	$P_{1a}=.3$
$y_{1b}=\ 80$	$\pi_b=.4$		$P_{1b}=.5$
			$W_0=200$

$=9.5370$) is worked out in Table 2 (column (4)).

Now, suppose that a single individual at time 0 secures *private*, *prior*, and *sure* information that state a will obtain at time 1. Since one individual's choices would only negligibly affect the ruling prices, he could realize essentially all the market value $P_{1b}y_{1b}$ of his c_{1b} endowment (which he alone knows to be worthless) for reallocation to the purchase of more c_0 and/or c_{1a}. In the numerical example, this amount is $P_{1b}y_{1b}=.5(80)=40$. With the parameters assumed, it can be shown[10] that he will purchase just $40/P_{1a}=133.3$ units of c_{1a} to add to his endowed 200 units (Table 2, columns (2) and (5)). If instead he were to learn that state b will obtain, he would reallocate the entire value $P_{1a}y_{1a}=60$ of his c_{1a} endowment to purchase $60/P_{1b}=120$ units of c_{1b}. Table 2 also shows the expected utility given perfect information (and the consequent rearrangement of consumption) to be substantially higher than the expected utility under uncertainty. These expectations are calculated, of course, using the individual's prior probability estimates as to what the information will reveal.

We now come to the crucial contrast. What of the *social* value of the sure information just analyzed above? Suppose that by a collective payment to some knowledgeable outsider, an entire community consisting of the representative individuals above could all simultaneously be informed as to which future state will obtain—how large a payment would they be justified in making? The answer is: None at all! Such information would be absolutely valueless to the community as a whole. Information is of value only if it can affect action. But with identical endowments, preferences, and beliefs in a world of pure exchange, all individuals must still end up holding their endowment time-state distributions. The only thing that could happen, given the new public

TABLE 2—PRIVATE VALUE OF INFORMATION

	Consumptive Choices[a]			Utility[b]		
	Uncertainty[c] (1)	State a[d] to obtain (2)	State b[e] to obtain (3)	Uncertainty (4)	State a to obtain (5)	State b to obtain (6)
c_0	100	100	100	4.6052	4.6052	4.6052
c_{1a}	200	333.3	—	.6 (5.2983)	5.8091	—
c_{1b}	80	—	200	.4 (4.3821)	—	5.2983
Expected Utility Under Uncertainty:				9.5370		
Conditional Utility:					10.4143	9.9035
Expected Utility Given Perfect Information:[f]					10.2100	

[a] $P_{1a}=.3$, $P_{1b}=.5$, $W_0=200$

[b] Computed according to: $U=log_e c_0+\pi_a log_e c_{1a}+\pi_b log_e c_{1b}$

[c] $\pi_a=.6$, $\pi_b=.4$

[d] $\pi_a=1$, $\pi_b=0$

[e] $\pi_a=0$, $\pi_b=1$

[f] Using prior weights $\pi_a=.6$, $\pi_b=.4$

---

[10] See Hirshleifer (1970b, pp. 9–11).

TABLE 3—SOCIAL VALUE OF INFORMATION

	Consumptive Choices			Utility		
	Uncertainty[a] (1)	State a[b] to obtain (2)	State b[c] to obtain (3)	Uncertainty (4)	State a to obtain (5)	State b to obtain (6)
c_0	100	100	100	4.6052	4.6052	4.6052
c_{1a}	200	200	—	.6 (5.2983)	5.2983	—
c_{1b}	80	—	80	.4 (4.3821)	—	4.3821
Expected Utility Under Uncertainty:				9.5370		
Conditional Utility:					9.9035	8.9873
Expected Utility Given Perfect Information:[d]					9.5370	

[a] $\pi_a=.6$, $\pi_b=.4$; $W_0=200$; $P_{1a}=.3$, $P_{1b}=.5$
[b] $\pi_a=1$, $\pi_b=0$; $W_0=200$; $P_{1a}=.5$, $P_{1b}=0$
[c] $\pi_a=0$, $\pi_b=1$; $W_0=200$; $P_{1a}=0$, $P_{1b}=1.25$
[d] Using prior weights $\pi_a=.6$, $\pi_b=.4$.

information, is that prices shift immediately to permit "sustaining" the endowment vector in the face of the changed beliefs entering into the utility function. In the numerical example, sure public information that state a will obtain, available *prior* to the decisions at $t=0$, will cause P_{1a} to rise to .5 (while P_{1b}, of course, falls to zero). Sure public prior information that state b will obtain raises P_{1b} to 1.25, while P_{1a} falls to zero. Table 3 confirms that in these circumstances the individual will choose the same consumptions with the same probabilities as in the original uncertainty situation.

One other very important consideration must now be taken into account. There is a possibility of still greater gain for the *privately* informed individual if he is permitted to speculate rather than merely move directly to his preferred consumptive position. Assuming private knowledge that state a was to obtain, for example, an optimally speculating individual would first convert not just his c_{1b} holdings but rather *all* of his wealth $W_0=200$ to c_{1a} holdings at the old price relationships. The anticipation here is that the true information will become public, P_{1a} rising to .5 and wealth to 333.3, prior to the finalizing of the consumption-investment decisions at $t=0$. Note that the individual with private information would have every incentive to publicize that information, *after* making his speculative commitment. The enormously enhanced private results achievable via speculation are detailed in Table 4.

So far, two reaction modes of the privately informed individual have been considered: consumptive adaptation, and optimal speculation. A third and even more attractive possibility is *resale of the information* itself. The potential gain then becomes enormous, since the individual is no longer constrained by his personal commodity endowment. However, it may not be easy for an informed individual to authenticate possession of valuable foreknowledge for resale purposes. After all, anyone could *claim* to have such knowledge. Feasible and optimal resale strategies, and the market value of resold information, are issues that cannot be pursued here. The crucial point remains that *the community as a whole obtains no benefit, under pure exchange, from either the acquisition or the dissemination* (*by resale or otherwise*) *of private foreknowledge.*

The contrast between the private profitability and the social uselessness of foreknowledge may seem surprising. In-

TABLE 4—PRIVATE VALUE OF INFORMATION, WITH SPECULATION

	Consumptive Choices			Utility		
	Uncertainty[a]	State a[b] to obtain	State b[c] to obtain	Uncertainty	State a to obtain	State b to obtain
c_0	100	166.7	250	4.6052	5.1160	5.5215
c_{1a}	200	333.3	—	.6 (5.2983)	5.8091	—
c_{1b}	80	—	200	.4 (4.3821)	—	5.2983
Expected Utility Under Uncertainty:				9.5370		
Conditional Utility:					10.9251	10.8198
Expected Utility Given Perfect Information:[d]					10.8830	

[a] $\pi_a=.6$, $\pi_b=.4$; $W_0=200$; $P_{1a}=.3$, $P_{1b}=.5$

[b] $\pi_a=1$, $\pi_b=0$; $W_0=333.3$; $P_{1a}=.5$, $P_{1b}=0$

[c] $\pi_a=0$, $\pi_b=1$; $W_0=500$; $P_{1a}=0$, $P_{1b}=1.25$

[d] Using prior weights $\pi_a=.6$, $\pi_b=.4$.

formation is widely considered to be a classic example of a "collective good," the type of commodity for which private incentives are supposed to lead to underprovision rather than over-provision on the market. Indeed, there may be something of a collective-good aspect to the *market* information alluded to earlier: information that helps improve an otherwise imperfect process of exchange. But the expenditure of real resources for the production of *technological* information is socially wasteful in pure exchange, as the expenditure of resources for an increase in the quantity of money by mining gold is wasteful, and for essentially the same reason. Just as a smaller quantity of money serves monetary functions as well as a larger, the price level adjusting correspondingly, so a larger amount of foreknowledge serves no social purpose under pure exchange that the smaller amount did not.

IV. The Value of Foreknowledge: Production and Exchange

Consider now the more realistic regime in which production and exchange both take place. Assume that endowments are just the same as before, for all individuals. But suppose that, in addition, each representative individual has a small discrete productive investment opportunity of the following form: exactly 1 unit of endowed c_0 may be sacrificed to produce additional income in either time-state $1a$ or time-state $1b$ (but not both). Suppose that the choice is between a yield of $2\frac{1}{2}$ units in state a or $2\frac{1}{2}$ units in state b. With the prices of the initial situation in the example above ($P_{1a}=.3$, $P_{1b}=.5$), every representative individual would choose the latter alternative; he would physically invest, transforming his endowment combination ($y_0=100$, $y_{1a}=200$, $y_{1b}=80$) into the attained combination ($c_0=99$, $c_0=200$, $c_{1b}=82.5$). Since the scale of the collective investment is not infinitesimal, the prices change slightly but not by enough to modify the desirability of the selection made.

Suppose now that *one single* individual is given sure, prior, and private information that state a will obtain. Here it would be socially desirable that this individual's investment sacrifice of c_0 (and everyone else's, as well) be redirected so as to produce c_{1a} instead of the useless c_{1b}. But if the information is private, the original prices must still be ruling so that the individual's incentives for production decisions remain unchanged. He will continue to invest for a c_{1b} return, despite knowing that the latter will turn out to be valueless. It is more profitable for him to commit the resources to c_{1b}, merely taking care to arrange in advance for the liquidation of the $2\frac{1}{2}$ units of c_{1b} (in addition, of course, to his endowed 80 units) at the currently ruling market prices. *Thus, as under the regime of pure exchange, private foreknowledge makes possible large private profit without leading to socially useful activity.* The individual would have just as much incentive as under pure exchange (even more, in fact) to expend real resources in generating socially useless private information.

What of the value of *public* information? If the information were made public that state a would obtain, P_{1a} would jump to .5 (P_{1b} falling to zero). Then the individual investments would all be shifted so as to yield c_{1a} instead of c_{1b}. This, of course, is socially efficient behavior. *Public information as to which state will obtain is indeed of social value in a regime of production and exchange.* However, it remains true that the value of private foreknowledge is enormously greater to any individual than the value to him of public foreknowledge. In the example used here, public information enables the representative individual to attain the consumption sequence ($c_0 = 99$, $c_1 = 202.5$) with probability .6, or the sequence ($c_0 = 99$, $c_1 = 82.5$) with probability .4. Private information enables him to attain with probability .6 the sequence ($c_0 = 99$, $c_1 = 337.5$)—based on converting his 82.5 units of c_{1b} at the original price ratio into 137.5 units of c_{1a} to be added to his endowed 200 units—or with probability .4 the sequence ($c_0 = 99$, $c_1 = 202.5$). Evidently, the possibilities with private information are far superior (still leaving aside the prospect of much greater gains through speculation and/or resale). Thus, the incentives for the use of resources to generate private information remain excessive.

What about speculation and resale? Having undertaken a speculative commitment, it is in the interest of the informed individual to publicize the information. Whether or not involved in speculation, the informed individual would find it advantageous wherever possible to resell the information. Under pure exchange, where foreknowledge is socially valueless, devotion of resources to dissemination is only a further social waste. In a regime of production, however, universal dissemination would improve the choice of investments. The partial dissemination that would ensue from private publicizing or through resale would also tend to shift prices and lead to some productive adaptation. Thus, *in a regime of production, the dissemination of information has social utility,* against which gain must be offset, of course, any resource cost of the dissemination process.

V. Distributive Considerations, Public Information, and Homogeneity of Beliefs

The key factor underlying all the results obtained above is the distributive significance of private foreknowledge. When private information fails to lead to improved productive alignments (as must necessarily be the case in a world of pure exchange, and also in a regime of production unless there is dissemination effected in the interests of speculation or resale), it is evident that the individual's source of gain can only be at the expense of his fellows. But even where information is disseminated and does lead to improved productive commitments, the distributive transfer gain will surely be far greater than the relatively minor productive gain the individual might reap from the redirection of his own real investment commitments.

Will *public* information have distributive implications, and if so, will this consideration provide additional private motivation for the generation of public information? The nature and direction of possible distribution effects turn upon the timing of information acquisition in comparison with the schedule of trading. Two alternative timing patterns will be considered here: 1) The information is publicly released before the opening of trading, or 2) the individuals trade to their consumptive optimum positions prior to the release of the information, with another round of trading permitted afterward.[11] Both trading and generation of information are best regarded as essentially continuous interacting processes, so that the second timing pattern seems more acceptable.[12] But the first is useful in emphasizing certain aspects of reactions to public information.

We will therefore imagine, first, a situation of pure exchange in which the true state of the world is announced in advance of any market trading whatsoever—while individuals are still at their endowment positions. Announcement that state a will obtain (or, more generally, release of any information tending to increase the probability π_a that individuals attach to state a) will, of course, enhance the position of those disproportionately endowed with state a claims. The rise in the price P_{1a} will enrich such individuals. It will also enhance the position of those whose tastes or beliefs previously inclined them in the direction of purchase of state b claims. For, such individuals would otherwise have largely wasted their income endowments in the purchase of worthless state b claims.

That public information has distributive implications does not, however, lead necessarily to the conclusion that private individuals will want to generate public information. Individuals disproportionately endowed with state a claims, for example, cannot be sure in advance that the information will not point to state b rather than state a. And indeed, it can be shown that in the circumstances assumed here, risk-averse individuals will prefer that the information not be released.[13] For, the anticipation of public information becoming available in advance of trading adds a significant *distributive risk* to the underlying *technological risk* (as to which state will obtain). A community of such individuals would actually pay something to a knowledgeable outsider not to reveal, in advance of market trading, which state will obtain! (This conclusion would have to be modified somewhat under a regime of production and exchange; the gain from redirection of productive investments achieved in consequence of the public information would have to be offset against the increased distributive risk.)

Let us turn now to the more reasonable assumption that individuals have already optimally adapted their decisions to their opportunities prior to the release of new public information. Differences of endowments would no longer have any relevance, and so there would be no net incentive for or against the acquisition of socially neutral information under pure exchange. (And there would be an appropriately small incentive for any one individual to support the acquisition of beneficial public information in a world of production.) A similar argument can be made about differences in tastes or beliefs so long as the individuals may be presumed to have merely moved to their consumptive optimum positions, and not engaged in speculative behavior. But, we know that for individuals with superior information there is a strong incentive to take speculative positions. Now, differences of beliefs amount to the same thing as *each individual's thinking that he is in possession of superior information.* Such differences open up a new range of possibilities.

---

[11] If no trading were permitted once the information was revealed, individuals' consumptive baskets (here, holdings of c_0 and c_1) would in general be non-Pareto optimal. That is, differences across individuals in marginal rates of substitution between commodities would persist. Such a model has been studied by Ross Starr. The non-Pareto optimality stemming from informational differences has also been noted by Earl A. Thompson (fn. 5) and by Arrow (1969, pp. 54–56).

[12] The second timing pattern corresponds to the dictum that, at any moment of time, the market has already "discounted" (allowed for) all publicly available information.

[13] See Hirshleifer (1970b, pp. 20–22).

We saw above that it was privately rational, for a better informed individual, to expend resources for the dissemination of socially neutral information—after having adopted a speculative commitment. With inhomogeneity of beliefs every person may be better informed, in his own opinion! Thus, the generation of public information is (from his point of view) nothing other than the dissemination of information already privately available to him. He will expect to reap speculative profits from this process. But so will other individuals, with quite opposed opinions! We therefore have rather strong grounds to anticipate that in these circumstances excessive resources will be devoted to the generation of public information.

These considerations may be clarified by reference to a well-known activity for the generation of public information—horse racing. Viewed as a research activity, horse racing may be presumed to have a small positive social value: the identification of faster horses works "to improve the breed." This consideration is evidently a very minor motivating factor for the activity in comparison with the opportunity to speculate upon one's supposedly superior knowledge. Without differences of opinion, it is said, there would be no horse races. That is, the social value is insufficient to motivate the research—the activity is founded upon the contradictory expectations of speculative gain.

Suppose that it costs $100 in real resources to run a horse race, and that the social advantage of knowing which is the fastest horse is just $5. Evidently, if the race is run society is engaging in excessive research. Now imagine that the potential speculative gain, to an individual convinced that his horse is truly faster, is just $90—he could still not earn enough, himself, to cover the costs of the race. But if several individuals are so convinced, each about his own horse, they may cooperate to stage the experiment. So conflict of beliefs may enormously compound the speculative factor that, even from the point of view of a single individual, tends to promote excessive investment in information-generating activity.

VI. The Value of Discovery Information

The acquisition of technological information usually refers to the detection of properties of Nature that permit the development of new tools or the utilization of new techniques. This is the type of information categorized as discovery above, in which Nature's secret will not be autonomously revealed but must be extracted by man. The necessity for human interposition makes the analysis of the value of discovery information somewhat more complex than the analysis of foreknowledge information.

For concreteness, consider the following situation. Suppose that if an alloy with an enormously high melting point of X° can be created, extremely cheap thermonuclear power will become feasible. The underlying state of the world is not the result of a probabilistic process: such an alloy may in fact be possible to create (state a) or may not (state b). While this is not a stochastic situation, it has been shown to be useful even in such circumstances to summarize our uncertainty in the form of a probability distribution.[14] Thus, just as in the case of foreknowledge, we can assume that individuals assign probabilities π_a and π_b to the two underlying states of the world. However, even if the favorable state a is the true one, Nature is not going to tell us herself. In the discovery situation, no news is bad news.

---

[14] See Leonard J. Savage.

From the point of view of any individual, however, the picture may not look very different from that analyzed under the heading of foreknowledge. For any individual there is a certain probability π_A of "good news" (discovery of the alloy), due to the actions of other men if not of his own. We can think of a compound event A which consists of the joint happenings "State a is true (the required alloy is possible) *and* this fact is successfully exploited (the alloy is created) within the time-period envisaged." Evidently, $\pi_A \leq \pi_a$, the probability of good news is generally less than the probability attached to the more favorable state of Nature. And similarly, if the event B is defined in a complementary way as representing "no news" (identical with "bad news" in the circumstances considered), we have $\pi_B \geq \pi_b$. The individual's decisions—for example, whether to invest in a productive process whose profitability will be highly sensitive to the prospect of cheap thermonuclear power—will run in terms of the probabilities of good and bad news rather than the probabilities attached to the states of nature.

With this modification, the analysis is essentially similar to what has gone before. We need only consider the more general regime of production and exchange. Given private, prior, and sure information of event A, the individual in a world of perfect markets would *not* adapt his productive decisions if he were sure the information would remain private until after the close of trading. For, prices of the time-state claims c_{1A} and c_{1B} (involving the observable states A and B, of course, not the "natural" states a and b which do not directly affect markets) will not have changed. However, as before it would be in his interest to speculate and/or resell the information, in which case prices will tend to shift. The obvious way of acquiring the private information in question is, of course, by performing technological research.[15] By a now familiar argument we can show once again that the distributive advantage of private information provides an incentive for information-generating activity that may quite possibly be in excess of the social value of the information.

The conclusions reached in the analysis of foreknowledge with respect to public information again carry over to the discovery situation. Public information is socially valuable in redirecting productive decisions, and to that extent individuals will rationally combine (through government and other instruments) to generate public information (or, in some cases, it will pay even a single individual to do so). But disparities of beliefs (differences in probability estimates) may lead to agreement upon a procedure of generating public information where social costs exceed the social benefit.

VII. Implications for Patent Policy

Eli Whitney obtained one of the first American patents, in 1794, for his cotton gin. With some business associates, he spent many years and invested considerable resources in the attempt to protect his patent and prosecute infringements. These efforts were largely fruitless. It is reasonable to infer that potential inventors, both before and after Whitney, have been deterred from searching for new knowedge by the fear of a similar outcome—hence the argument for effective patent protection. On the other hand, had

---

[15] This is not the only way. The information might be purchased (or stolen) from some other person.

Whitney succeeded in obtaining the terms he demanded from users of his idea, the enormous expansion that actually took place in the production and consumption of cotton would have been significantly hampered.[16] This conflict between the "static" disadvantage of a patent monopoly and the "dynamic" advantage of encouraging invention is quite properly emphasized in the traditional literature.[17]

But what seems to have been overlooked is that there were other routes to profit for Whitney. The cotton gin had obvious speculative implications for the price of cotton, the value of slaves and of cotton-bearing land, the business prospects of firms engaged in cotton ware-housing and shipping, the site values of key points in the transport network that sprang up. There were also predictable implications for competitor industries (wool) and complementary ones (textiles, machinery). It seems very likely that some forethoughted individuals reaped speculative gains on these developments, though apparently Whitney did not. And yet, he was the first in the know, the possessor of an unparalleled opportunity for speculative profit. Alternatively, of course, Whitney could have attempted to keep his process secret except to those who bought the information from him.

The issues involved may be clarified by distinguishing the "technological" and "pecuniary" effects of invention. The technological effects are the improvements in production functions—interpreted in the widest sense to include the possible production of new commodities, the discovery of new resources, etc.—consequent upon the new idea. The pecuniary effects are the wealth shifts due to the price revaluations that take place upon release and/or utilization of the information. The pecuniary effects are purely redistributive.[18]

For concreteness, we can think in terms of a simple cost-reducing innovation. The technological benefit to society is, roughly, the integrated area between the old and new marginal-cost curves for the preinvention level of output plus, for any additional output, the area between the demand curve and the new marginal-cost curve. The holder of a (perpetual) patent could ideally extract, via a perfectly discriminatory fee policy, this entire technological benefit. Equivalence between the social and private benefits of innovation would thus induce the optimal amount of private inventive activity. Presumably, it is reasoning of this sort that underlies the economic case for patent protection. It is true that under a patent system there will, in general, be some shortfall in the return to the inventor, due to costs and risks in acquiring and enforcing his rights, their limited duration in time, and the infeasibility of a perfectly discriminatory fee policy. On the other side are the recognized disadvantages of patents: the social costs of the administrative-judicial process, the possible anti-competitive impact, and restriction of output due to the marginal burden of patent fees.[19] As a second best kind of judgment, some degree of patent protection has seemed a reasonable compromise among the objectives sought.

But recognition of the unique position of the innovator for forecasting and consequently capturing portions of the *pecuniary* effects—the wealth transfers due to price revaluations—may put matters in a different light. The "ideal" case of the perfectly discriminating patent holder earning the entire technological benefit is no longer so ideal. For, the same inventor is in a position to reap speculative profits, too; counting these as well, he would clearly be overcompensated.

---

[16] Whitney and his partners planned to retain all the ginning in their own hands, buying the raw product and selling the ginned cotton (Dumas Malone, vol. 10, p. 159).

[17] See Machlup.

[18] For a discussion in the context of government resource-investment policy, see Roland McKean (ch. 8).

[19] A perfectly discriminatory fee system would place no marginal burden and thus would not lead to any restriction of output.

Consider now the opposite extreme. Do we have reason to believe that the potential speculative profits to the inventor, from the pecuniary effects that will follow release of the information at his unique disposal, will be so great that society need take no care to reserve for him any portion of the technological benefit of his innovation? The answer here is indeterminate. There is no logically necessary tie between the size of the technological benefit on the one hand, and the amplitude of the price shifts that create speculative opportunities on the other.[20]

Even if the prospective price revaluations are ample, however, there will be limitations to the inventor's capacity to profit from them. For, speculative profits are constrained by the magnitude of feasible speculative commitments. If the possessor of prior information acts alone, he is limited by what may be a puny wealth endowment. But if he tries to sell his information, in effect buying a share in a larger speculative pool, he will find it difficult to consummate such a transaction. The most important limitation of all has not heretofore been taken up in this paper, but must be considered in a policy discussion: imperfection of markets for time-state claims.[21] Given the inconceivably vast number of potential contingencies and the costs of establishing markets, the prospective speculator will find it costly or even impossible to purchase neutrality from "irrelevant" risks. Eli Whitney could not be *sure* that his gin would make cotton prices fall: while a considerable force would clearly be acting in that direction, a multitude of other contingencies might also have possibly affected the price of cotton. Such "uninsurable" risks gravely limit the speculation feasible with any degree of prudence.

We are left, therefore, in an agnostic position. The fundamental argument for patent protection is gravely weakened when it is recognized that the pecuniary effects of the invention are a potentially enormous source of return to the inventor, quite apart from the technological benefit that the patent system attempts to reserve for him. But we cannot show that no patent protection at all is warranted, that the profits from speculation or from resale[22] suffice for an appropriate inducement to invention. These profits may more than suffice, or they may fall substantially short; there is no necessary relation at all between the magnitudes of the technological and the pecuniary effects. Or, more precisely, between the magnitudes of that fraction of the technological effect that a patentee can capture and that fraction of the pecuniary effect that a speculator on prior information can capture.

VIII. Summary

In the model of this paper, markets are assumed to be perfect. Uncertainty attaches only to individuals' perceptions of their endowments and productive opportunities (technological uncertainty). The private and social values of two main categories of technological information were considered: 1) foreknowledge of states of the world that will in time be revealed by Nature herself (e.g., the weather), and 2) discovery of hidden properties of Nature that can only be laid bare by human action.

---

[20] A relatively minor shift in locomotive technology, for example, might lead railroad planners to select an entirely different route for a new line, with drastic upward and downward shifts of land values. Paul Samuelson emphasizes, p. 974, the disproportionality between the gain reaped by the first-in-time speculator and the social utility of his activity.

[21] This imperfection has been emphasized by Arrow (1962, and also 1969).

[22] Resale of information does not stand on quite the same footing as speculation. Speculation is an extra source of gain, whether or not patent has been obtained, whereas resale of information otherwise kept secret is an *alternative* to obtaining a patent.

Private information that remains private was shown to be of no social value—in the sense of being purely redistributive, not leading to any improvement in productive arrangements. There is an incentive for individuals to expend resources in a socially wasteful way in the generation of such information. Public information, in contrast, does affect productive decisions in a socially appropriate way. Speculative profits from the price revaluations to be anticipated provide the knowledgeable individual with an incentive to disseminate (publicize) his private information. Still greater profit is possible if the information can be resold. In a world of pure exchange, there will in general be private overinvestment in information: resources committed to acquisition and to dissemination are both wasted from the social point of view. In a world of production, however, the gains from productive rearrangements due to the information must be offset against the costs of acquisition and dissemination; there may or may not be private overinvestment.

Distributive considerations enter also into the motivation for the acquisition of *public* information. To the extent that the prospect of such information imposes a distributive risk upon individuals—due, for example, to possible revaluation of endowment holdings—there will be an aversion to socially neutral and even (to some degree) to socially beneficial public information. Probably more important is a force acting in the opposed direction. With inhomogeneous beliefs, individuals with differing opinions will tend each to believe that revelation of new information will favor his own speculative commitments. Hence, a group of such individuals might willingly cooperate in making expenditures far in excess of the social value of the information to be acquired.

The standard literature on the economics of research and invention argues that there tends to be private underinvestment in inventive activity, due mainly to the imperfect appropriability of knowledge. The contention made is that, even with a patent system, the inventor can only hope to capture some fraction of the technological benefits due to his discovery. This literature overlooks the consideration that there will be, aside from the technological benefits, pecuniary effects (wealth redistributions due to price revaluations) from the release of the new information. The innovator, first in the field with the information, is able through speculation or resale of the information to capture a portion of these pecuniary effects. This fact is socially useful in motivating release of the information. Even though practical considerations limit the effective scale and consequent impact of speculation and/or resale, the gains thus achievable eliminate any a priori anticipation of underinvestment in the generation of new technological knowledge.

REFERENCES

A. Alchian, "Information Costs, Pricing, and Resource Unemployment," in E. S. Phelps et al., eds., *Microeconomic Foundations of Employment and Inflation Theory*, New York 1970.

K. J. Arrow, "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Economic Activity: Economic and Social Factors*, Universities–Nat. Bur. Econ. Res. Conference series, Princeton 1962.

———, "Le Rôle des Valeurs Boursières pour la Repartition la Meilleure des Risques," *International Colloquium on Econometrics, 1952*, Centre National de la Recherche Scientifique, Paris 1953; translated as "The Role of Securities in the Optimal Allocation of Risk-Bearing," *Rev. Econ. Stud.*, Apr. 1964, *31*, 91–96.

———, "Discounting and Public Investment Criteria" in A. V. Kneese and S. C. Smith, eds., *Water Research*, Baltimore 1966.

———, "The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Nonmarket Allocation," in U.S. Congress, Joint Economic Committee, *The Analysis and Evaluation of Public Expenditures: The PPB System*, a compendium of papers submitted to the Subcommittee on Economy in Government, 91st Cong., 1st sess., Washington 1968, *1*, 47–64.

K. Borch, "Equilibrium in a Reinsurance Market," *Econometrica*, July 1962, *30*, 424-44.

G. Debreu, *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, New York, 1959.

H. Demsetz, "Information and Efficiency: Another Viewpoint," *J. Law Econ.*, Apr. 1969, *12*, 1–22.

P. A. Diamond, "The Role of a Stock Market in a General Equilibrium Model with Technological Uncertainty," *Amer. Econ. Rev.*, Sept. 1967, *57*, 759–76.

J. H. Drèze and F. Modigliani, "Epargne et Consommation en Avenir Aléatoire," *Cahiers du Seminaire d'Econometrie*, 1966, *9*, 7–33.

E. F. Fama and A. B. Laffer, "Information and Capital Markets," *J. Bus. Univ. Chicago*, July 1971, *44*.

J. Hirshleifer, "Investment Decision under Uncertainty: Choice-Theoretic Approaches," *Quart. J. Econ.*, Nov. 1965, *79*, 509–36.

———, "Investment Decision under Uncertainty: Applications of the State-Preference Approach," *Quart. J. Econ.*, May 1966, *80*, 252–77.

———, (1970a) *Investment, Interest, and Capital*, Englewood Cliffs 1970.

———, (1970b) "The Private and Social Value of Information and the Reward to Inventive Activity," working paper no. 158, Western Management Science Institute, UCLA, Apr. 1970.

T. Koopmans, *Three Essays on the State of Economic Science*, New York 1957.

F. Machlup, "Patents," *International Encyclopedia of the Social Sciences*, New York 1968, *11*, 461–72.

D. Malone, *Dictionary of American Biography*, New York 1964.

H. G. Manne, *Insider Trading and the Stock Market*, New York 1966.

J. Marschak, (1968a) "Decision-making: Economic Aspects," *International Encyclopedia of the Social Sciences*, New York 1968, *4*, 42–55.

———, (1968b) "Economics of Inquiring, Communicating, Deciding," *Amer. Econ. Rev. Proc.*, May 1968, *58*, 1–18.

R. N. McKean, *Efficiency in Government through Systems Analysis*, New York 1958.

S. A. Ozga, "Imperfect Markets through Lack of Knowledge," *Quart. J. Econ.*, Feb. 1960, *74*, 29–52.

P. A. Samuelson, "Intertemporal Price Equilibrium: A Prologue to the Theory of Speculation," *Welwirtschaftliches Archiv*, Dec. 1957, *79*, 181–219, reprinted in J. E. Stiglitz, ed., *The Collected Scientific Papers of Paul A. Samuelson*, Cambridge, Mass. 1966, *2*, 946–84.

L. J. Savage, *The Foundations of Statistics*, New York 1954.

R. M. Starr, "The Effect of Individual Beliefs on the Achievement of Pareto Optimal Distribution under Uncertainty," working paper no. 127, Western Management Science Institute, UCLA, Oct. 1967.

G. Stigler, "The Economics of Information," *J. Polit. Econ.*, June 1961, *69*, 213–25.

———, "Information in the Labor Market," *J. Polit. Econ.*, Oct. 1962 supp., *70*, 94–105.

E. A. Thompson, "A Pareto Optimal Group Decision Process," in G. Tullock, ed., *Papers on Non-Market Decision Making*, Charlottesville 1966.

Entrepreneurial Income: How Does It Measure Up?

By LAURITS R. CHRISTENSEN*

A long standing problem in *U.S.* social accounting has been the apparent impossibility of reconciling the theory of competitive factor pricing with the observed income distribution. Studies of income distribution have invariably found that the income of entrepreneurs in the United States has been so small as to preclude the interpretation that the entrepreneur has received a competitive return to both his labor and his capital. This conclusion is well accepted. For example, Edward F. Denison (1967, p. 39) recently asserted: "Proprietors' income in the United States is not large enough to accommodate both the assumption that the rate of return on capital and land is the same as in corporations and the assumption that the average earnings of proprietors and unpaid family workers are the same as those of employees."

The purpose of this paper is to reexamine *U.S.* entrepreneurial income for the 1929–69 period. The basic division of entrepreneurial income is achieved by imputing a competitive wage to self-employed workers and allocating the residual income to capital. The implied rate of return on noncorporate capital is then computed and compared to the rate of return on corporate capital. This methodology is not novel, but the conclusions reached are markedly different from those of previous studies, which found that the noncorporate rate of return is much smaller than the corporate rate of return. Unfortunately, the rates of return used for comparison have contained serious flaws. The comparison is meaningful only if rates of return which are relevant to the investor are used. This of course implies the use of after-tax rates of return. Previous studies, however, have compared before-tax rates of return. A second major flaw has been the exclusion of capital gains from the compared rates of return. The total return to an asset consists of income on current account and capital gains; there does not seem to be any justification for excluding either part in computing the rate of return on the asset.

When the appropriate rates of return are compared, the alleged dominance of the corporate rate of return over the noncorporate rate is not substantiated. The average after-tax rates of return over the full sample period are nearly equal—7.6 percent for the noncorporate sector versus 7.3 percent for the corporate sector. Considering the difficult measurement probems, this difference is insignificant. Moreover each rate of return exceeds the other in approximately half of the years. We conclude that the evidence supports the proposition that noncorporate and corporate rates of return tend to equality. The observed distribution of income for the United States is consistent with this prediction of the theory of competitive factor pricing.

---

* Assistant professor of economics at the University of Wisconsin. I wish to acknowledge research support from the National Science Foundation and the University of Wisconsin Graduate School; and expert research assistance by Victoria Farrell. This paper was presented at the Second World Congress of the Econometric Society, Cambridge, England, September 7–14, 1970. Edward Denison, Dale Jorgenson, Robert Krainer, Jeffrey Williamson, and a referee provided helpful comments on previous drafts, but I retain sole responsibility for any shortcomings of the final product.

I. Previous Studies of Entrepreneurial Income

Numerous studies have included an analysis of entrepreneurial income. The usual motivation has been to divide entrepreneurial income into labor and property components in order to achieve an exclusive division of national income into property and labor income. Probably the most familiar of such studies is that of Irving B. Kravis. Even though Kravis acknowledged that his procedures could bear considerable refinement, his conclusions are consistent with previous findings and well accepted by subsequent researchers. Thus it may be useful to summarize its relevant parts. Kravis uses four procedures to allocate entrepreneurial income to capital and labor input. The first procedure employs the assumption that "the rate of yield on entrepreneurial property has been the same as that upon other property" (p. 924). The remaining income is then allocated to entrepreneurial labor. The second procedure assumes that entrepreneurial labor receives the same return as other labor and allocates the residual income as return to property. These two procedures produce widely divergent results. The first procedure, termed the asset basis, allocates substantially more income to property than the labor basis. This holds for all figures that Kravis reports (including those of D. Gale Johnson) for 1900–57. The implication is that entrepreneurial income has not been sufficient to provide competitive returns to both labor and capital. If the discrepancy were slight, it could be easily ignored, but the large differences seem to demand a compromise. Kravis provides two possibilities, a proportional basis and an economy-wide basis. The proportional basis sets fixed shares of property and labor income out of entrepreneurial income—Kravis suggests 35 percent and 65 percent based on Johnson. The economy-wide basis computes the relative shares of property and labor in all other national income and uses the result to allocate entrepreneurial income to property and labor. These latter procedures entirely ignore economic forces in the noncorporate sector. The only redeeming features of these compromises are the "reasonable" distribution of income resulting, and the inherent simplicity of implementation. Writers following Kravis have found compromise solutions attractive. Stanley Lebergott says "Kravis' solution is surely in the right direction. . ."; Denison (1962, 1967) uses a version of Kravis' economy-wide basis, but prefers to link the entrepreneurial allocation to that observed in the corporate sector. The view that entrepreneurial income has not provided competitive returns to labor and capital input has gone unchallenged.

II. The Labor Share of Entrepreneurial Income

The opportunity cost of a self-employed worker's time is the wage rate which he could achieve doing similar work as an employee. An ideal measure of entrepreneurial labor income would consist of this wage times hours worked, computed for all self-employed workers in the economy. Unfortunately, this measure would be impossible to achieve in practice. A good approximation could be achieved by classifying workers into categories according to the type of work and personal characteristics such as age, sex, race, region, education, experience, and health. Then the average wage rate for employees in each category could be imputed to self-employed workers in the same category. This would be an ambitious project in itself and is not attempted for the present study. Rather, we rely on the approximation used by Kravis: we assume that in the private sector of the economy self-employed workers and employees, on average, have the same personal characteristics and work the same number of hours per year at similar kinds of work. In other words we assume that average annual earnings per employee are equal to average annual earnings per self-employed worker.

The Kravis imputation is accomplished using data published by the Office of Business Economics (*OBE*) 1966. The average annual wage for employees is computed as the compensation of employees divided by the number of full-time equivalent employees in all private industries. The *OBE* estimate of the number of self-employed workers includes "active proprietors of unincorporated enterprises devoting the major portion of their time to the business" (p. 202, note 1). Our estimate of entrepreneurial labor income is the implicit average annual wage times the number of proprietors.

There are several possible biases resulting from our labor imputation. The levels of educational attainment differ enormously among self-employed workers; if the mix differs from that of employees there will be a bais. The many doctors, dentists, lawyers, and other professionals who are self-employed might make one suspect that the average educational level for self-employed workers is higher. On the other hand, the largest class of self-employed workers are farmers, who have a relatively low average education level. Whether the extremes cancel out is a matter of conjecture. Simon Kuznets (p. 27) suggests that self-employed workers on average have lower skill levels than employees and therefore earnings should be imputed to their labor services at a lower wage rate. However, there does not seem to be much evidence in the literature on this issue one way or the other.

A clear bias results from the fact that we have not made an adjustment for unpaid family workers. The *OBE* does not include an estimate of unpaid family workers in their series "persons engaged in production." The reason given for this exclusion is "unresolved difficulties in . . . definition and measurement" (p. 202, note 1). The imputation of labor earnings to unpaid family workers would increase the proportion of entrepreneurial income attributed to labor input.

Leonard G. Rosenberg contends that the *OBE* estimate of proprietors overstates the actual labor input. Many proprietors (including farmers) have other jobs; others prefer to take more leisure. He concludes, "Some deflation of the number of proprietors to a full-time equivalent basis is certainly required" (p. 143). This would result in a lower total of labor income imputed to entrepreneurs.

These and other biases may or may not cancel out. Rather than make any arbitrary adjustments, we proceed on the assumption that our labor imputation is correct—keeping in mind that some error may be built in at this stage.

III. The Return to Corporate and Noncorporate Capital

Economic theory predicts that if other things are equal, owners of capital will invest where the expected return to their capital is the greatest. If there are diminishing returns to capital as a factor of production, rates of return will tend to equalize as capital flows from areas of low return to areas of high return. Of course, other things are not equal, and to make use of this theory of differential investment we must make adjustments. The most important consideration is undoubtedly an adjustment for disparate tax treatments afforded various types of assets. In particular, the corporate income tax has crucial importance in the present study. Income attributable to capital in the corporate sector is subject to this tax while similar income in the noncorporate sector is not. The statutory rate of this taxation has been near 50 percent for most of the postwar period. Obviously, there must be offsetting advantages of the corporate form of business organization. Noncorporate business forms become unwieldy with large size and have limited opportunities for diversification; thus the corporate form is practically necessary in order to achieve economies of scale. Corporations may earn higher rates of return before tax due to economies of scale; but after-tax rates of return can still be equal because of larger corporate tax burden.

The attractiveness of an asset does not depend entirely on the after-tax current gross return as measured by income. The capital gains potential of an asset may form an important part of the expected rate of return. This is especially true since capital gains receive generally preferential tax treatment.

The computation of after-tax rates of return including allowance for replacement and capital gains requires estimates of after-tax returns, the current value of assets, and the economic depreciation of these assets. The market value of assets is difficult to determine; we use replacement cost as the best available measure of the market value of the assets in the corporate and noncorporate sectors.

Gross capital income in the corporate sector includes income originating in corporate business, less compensation of employees, plus capital consumption allowances, plus property taxes. Gross capital income in the noncorporate business sector includes income originating in sole proprietorships and partnerships and other private business, less compensation of employees, less the imputation to entrepreneurial labor, less the portions of rental income to persons and net interest which are imputed to owner-occupied dwellings and institutional buildings, plus capital consumption allowances,[1] plus property taxes. These measures of capital income are gross of depreciation and all direct taxes, but net of indirect business taxes other than property taxes. Plausible arguments could be made for adjustments to these measures of capital income, but these measures are quite standard and avoid controversial issues.[2] The present treatment follows the *U.S.* national income accounts except for capital gains, depreciation, and property taxes. Accrued capital gains are not included in the official *U.S.* accounts; thus we must construct our own estimates. The capital gain on an asset during the year is computed as the stock of capital at the beginning of the period times the difference between the prices prevailing this period and last period.[3] The capital consumption allowances in the *U.S.* accounts are not appropriate estimates of economic depreciation because they are based on accounting practices used for tax purposes. We attempt to estimate depreciation reflecting actual declines in efficiency.

---

[1] Except for the imputation to entrepreneurial labor and property taxes all these data are taken directly from *OBE* (1966) and subsequent National Income issues of the *Survey of Current Business*. The *OBE* has kindly supplied an estimate of state and local property taxes paid by corporations. A portion of business motor vehicle license fees and other indirect business taxes has also been included in corporate property taxes. The remainder of state and local property taxes, business motor vehicle license fees, and other indirect business taxes constitute noncorporate property taxes.

[2] For example, Christensen and Jorgenson have included business transfer payments, which are primarily corporate gifts to nonprofit institutions, in corporate capital income. They also included the statistical discrepancy in noncorporate capital income. Since the statistical discrepancy has tended to be positive, these adjustments largely offset each other in a computation of the differential between corporate and noncorporate rates of return. With these changes in the definition of income the average after-tax rates of return are 7.9 percent and 7.8 percent compared to 7.6 percent and 7.3 percent, respectively, as shown in Table 1.

[3] The price indices used to compute capital gains are as follows: (1) structures, constant cost 2 deflator *OBE* (unpublished); (2) producers' durables, implicit deflator from the national income and product accounts *OBE* (1966); (3) nonfarm inventories, Bureau of Labor Statistics (*BLS*) wholesale price index for goods other than farm products and food; (4) farm inventories, *BLS* wholesale price index for farm products; (5) land, Goldsmith and Manvel. See Christensen and Jorgenson for details and discussion.

There have been numerous modifications in depreciation allowances for structures and equipment which reflect policy decisions to attempt to simulate investment. The modifications do not reflect perceived changes in the rate of decline in efficiency of durable goods. Unfortunately, there is very little objective evidence to guide the estimation of economic depreciation. Alternative estimates of capital stock are often reported based on several different assumed types of economic depreciation. Robert C. Wasson, John C. Musgrave, and Claudia Harkins report alternative estimates for the *OBE* based on straight line and double declining balance depreciation. We have elected to use the double declining balance assumption to estimate economic depreciation. The main reason for this choice is that double declining balance is easier to implement than the straight line assumption.

Estimates of capital stock in the corporate and noncorporate sectors are not available in the national accounts. Utilizing the double declining balance depreciation assumption, these capital stocks can be constructed by a perpetual inventory method. Following the basic approach of Raymond Goldsmith, investments in various types of assets are allocated between corporate and noncorporate business. These investment series along with capital stock benchmarks and estimated rates of replacement are sufficient to generate the required capital stock and depreciation series. The stock series can then be valued at replacement cost. The specific assumptions and data used in constructing these series are reported in detail by Christensen and Jorgenson. The estimated values of corporate and noncorporate capital stock are shown in the first two columns of Table 1.[4] They include structures, equipment, land and inventories.[5]

The net return to capital before taxation is capital income, less current replacement of capital goods, plus accrued capital gains. The before-tax rates of return, computed by dividing the net return to capital by the value of tangible assets at the end of the previous year, are given in the third and fourth columns of Table 1.

Accrued capital gains are an inportant part of the net return to capital, but they are sometimes omitted in rate of return calculations. In order to exhibit the importance of capital gains we present the rate of capital gains for each of the sectors in the fifth and sixth columns of Table 1. These rates are computed by dividing accrued capital gains by the value of tangible assets at the end of the previous year.

Property taxes are deductible as a business expense by both corporate and noncorporate business. The effective rate of property taxation is computed by dividing estimated property taxes by the value of tangible assets at the end of the previous year. The effective rates of property taxation are shown in the seventh and eighth columns of Table 1.

The remaining problem is the determination of direct taxes on the returns to corporate and noncorporate capital. Direct taxes which impinge on the returns to capital include federal, state, and local income taxes and corporate profits taxes, federal estate and gift taxes, and state and local death and gift taxes. Corporate profits taxes can unambiguously be assigned

---

[4] Table 1 contains five-year averages and the forty-one year average for the full sample period. The yearly observations for all columns are shown in Table 2.

[5] Owner-occupied housing and consumer durables are excluded. They are not usually considered to be included in entrepreneurial or noncorporate business activity, but rather constitute household business activity. The capital stock figures shown are for the end of the previous year.

TABLE 1

	(1)	(2)	(3)	(4)	(5)	(6)
	Capital stock (dollars in billions)		Rates of return before taxation in percent		Rates of capital gains in percent	
	Noncorp.	Corp.	Noncorp.	Corp.	Noncorp.	Corp.
1929–33	93.9	104.3	−2.9	.3	−6.1	−4.6
1934–38	80.4	86.0	7.4	10.8	3.0	2.9
1939–43	93.6	102.0	15.7	21.7	5.7	5.3
1944–48	138.5	148.8	23.0	25.1	9.6	8.6
1949–53	214.3	255.9	11.7	17.3	2.8	3.1
1954–58	260.4	353.8	11.5	15.5	4.0	3.0
1959–63	334.8	458.6	10.8	13.9	3.8	1.5
1964–68	439.1	607.0	13.1	17.7	5.1	3.1
1929–69	214.8	276.8	11.3	15.4	3.6	2.9

	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Property taxation, effective rates in percent		Corporate income taxation:		Direct personal taxation of property income:		Rates of return after taxation in percent	
	Noncorp.	Corp.	Total taxes (dollars in billions)	Effective rate in percent	Total taxes (dollars in billions)	Effective rate in percent	Noncorp.	Corp.
1929–33	2.4	1.8	.7	*	.9	*	−5.6	−2.6
1934–38	2.8	2.1	1.1	17	1.1	16	4.2	6.7
1939–43	2.7	2.0	7.5	32	2.4	9	11.8	11.7
1944–48	2.1	1.7	11.3	35	6.0	12	18.6	14.0
1949–53	1.9	1.5	18.0	46	7.4	17	8.3	7.4
1954–58	2.1	1.6	20.3	43	10.0	17	7.7	6.7
1959–63	2.3	1.7	24.1	43	13.8	20	6.7	5.5
1964–68	2.6	1.8	33.5	35	19.7	16	8.8	8.7
1929–69	2.4	1.8	15.2	*	8.2	*	7.6	7.3

* Some divisors are negative or negligible.

to the corporate return. Total corporate income taxes levied by federal, state, and local governments appear in the ninth column of Table 1. Dividing corporate income taxes by the net return to corporate capital before direct taxation but net of property taxes yields the effective rate of corporate income taxation, which appears in the tenth column of Table 1.

Direct taxes other than corporate income taxes result from personal levies on wages and salaries, entrepreneurial income, corporate dividends, realized capital gains on corporate shares and noncorporate tangible assets, interest from private and government securities, and profits and interest from holdings of foreign securities and tangible assets. These taxes can be allocated between labor and property income in accord with a method developed by Lenore Frane and Lawrence R. Klein. They describe their method of allocation as follows: "The fraction of taxes in each income bracket allocated to wage income is estimated by the ratio of wage and salary income to 'adjusted gross income' in each bracket. Total taxes allocated to wages are the sum of the taxes so allocated in all brackets" (p. 336). This method (with an adjustment for entrepreneurial labor income) has been applied to the *U.S.* data for 1929–58 by A. Ando and E. C. Brown. We have updated the data of Ando and Brown to 1969 in a way that closely approximates their procedures.[6]

---

[6] Personal income taxes on income from labor services, as estimated by Ando and Brown, is a remarkably stable proportion of total personal income tax receipts. The

TABLE 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Capital stock (dollars in billions)		Rates of return before taxation in percent		Rates of capital gains in percent		Property taxation, effective rates in percent		Corporate income taxation:		Direct personal taxation of property income:		Rates of return after taxation in percent	
	Noncorp.	Corp.	Noncorp.	Corp.	Noncorp.	Corp.	Noncorp.	Corp.	Total taxes (dollars in billions)	Effective rate in percent	Total taxes (dollars in billions)	Effective rate in percent	Noncorp.	Corp.
1929	105.6	112.9	7.3	10.4	−.3	.0	2.1	1.7	1.4	14	1.5	10	4.7	6.8
1930	107.2	116.4	−3.6	1.7	−8.4	−5.9	2.2	1.7	.8	*	1.3	*	−6.3	−1.3
1931	98.0	110.1	−9.9	−4.6	−12.6	−8.6	2.2	1.7	.5	*	.8	*	−12.4	−7.1
1932	85.3	97.6	−11.5	−7.3	−11.8	−8.6	2.6	1.9	.4	*	.5	*	−14.2	−9.9
1933	73.7	84.4	3.2	1.5	2.4	.1	2.8	2.0	.5	*	.5	*	.2	−1.4
1934	73.9	80.9	7.8	10.3	5.9	5.2	2.7	2.0	.7	11	.6	5	4.9	7.0
1935	76.6	83.3	8.4	9.3	3.4	2.3	2.7	2.0	1.0	16	.8	8	5.2	5.7
1936	80.0	84.0	8.5	11.2	3.8	1.5	2.8	2.1	1.4	18	1.1	9	5.2	6.7
1937	83.1	86.7	10.5	17.2	4.5	6.8	2.8	2.1	1.5	12	1.6	8	7.1	12.2
1938	88.5	95.0	1.7	5.9	−2.8	−1.4	2.7	2.0	1.0	28	1.4	49	−1.5	1.9
1939	85.8	92.0	4.2	8.7	−.7	−.6	2.8	2.1	1.4	24	1.1	15	1.2	4.3
1940	85.4	91.7	9.1	14.9	2.7	2.0	3.0	2.2	2.8	24	1.1	7	5.7	8.9
1941	88.7	96.2	19.0	25.6	9.0	7.8	2.9	2.1	7.6	34	1.5	5	15.3	14.8
1942	98.7	109.2	23.8	31.4	10.2	11.1	2.6	1.9	11.4	36	2.6	6	19.9	17.9
1943	109.2	120.8	22.3	28.1	7.2	6.4	2.4	1.8	14.1	44	5.8	14	17.1	12.6
1944	116.0	126.3	19.4	25.5	5.6	4.8	2.3	1.8	12.9	43	5.6	14	14.7	11.5
1945	122.2	129.9	18.5	18.4	4.8	1.7	2.2	1.7	10.7	49	6.7	19	13.1	6.8
1946	128.4	131.7	30.9	25.3	15.0	11.2	2.1	1.7	9.1	29	6.1	10	26.0	15.0
1947	149.8	158.2	28.2	34.2	16.6	18.8	2.0	1.6	11.3	22	5.8	7	24.4	23.6
1948	176.2	197.9	18.0	22.3	5.9	6.6	1.9	1.5	12.5	30	5.8	9	14.6	13.1
1949	192.9	222.2	6.4	12.5	−2.3	−.4	1.8	1.5	10.4	43	5.3	19	3.7	5.1
1950	191.2	224.9	17.8	19.8	8.3	4.0	2.0	1.6	17.8	43	6.9	12	13.9	9.1
1951	213.3	246.2	18.3	24.9	8.2	8.7	1.9	1.5	22.3	39	7.6	10	14.7	12.9
1952	235.5	284.8	9.0	15.5	.3	2.3	1.9	1.5	19.4	49	8.4	20	5.7	5.7
1953	238.6	301.5	7.0	13.7	−.5	1.0	2.0	1.5	20.3	55	8.6	26	3.7	4.0
1954	239.9	314.0	9.0	12.9	1.7	.9	2.1	1.6	17.7	50	8.6	22	5.4	4.4
1955	245.8	321.8	9.8	16.8	2.2	1.9	2.1	1.6	21.6	44	9.2	18	6.3	6.9
1956	255.0	342.6	13.0	19.4	5.7	5.8	2.2	1.7	21.7	36	10.4	14	9.3	9.7
1957	271.5	380.2	13.0	16.8	5.8	4.7	2.2	1.6	21.2	37	11.0	15	9.2	8.1
1958	289.8	410.6	12.6	11.8	4.9	1.5	2.1	1.6	19.0	46	10.7	18	8.6	4.5
1959	306.1	420.7	9.7	14.5	2.8	2.0	2.2	1.6	23.7	44	11.8	19	6.1	5.8
1960	317.4	441.5	10.5	12.8	3.9	1.1	2.3	1.7	23.0	47	13.2	22	6.4	4.6
1961	332.5	459.3	11.2	12.6	4.3	1.1	2.3	1.7	23.1	46	13.6	21	7.0	4.6
1962	349.2	474.0	11.6	14.7	4.4	1.8	2.4	1.8	24.2	39	14.8	19	7.4	6.3
1963	368.8	497.5	10.9	14.8	3.6	1.6	2.5	1.8	26.3	41	15.8	20	6.7	6.2
1964	387.3	520.9	11.2	16.2	4.0	2.0	2.5	1.8	28.3	38	15.6	17	7.2	7.4
1965	407.4	552.4	13.6	18.0	5.3	2.6	2.6	1.8	31.3	35	17.2	15	9.3	8.9
1966	436.6	594.9	14.0	19.2	5.4	3.5	2.6	1.8	34.3	33	19.6	15	9.6	9.9
1967	467.3	656.1	13.3	17.4	4.9	3.6	2.6	1.7	33.2	32	21.3	16	8.9	8.9
1968	497.1	710.6	13.5	17.6	5.6	3.8	2.6	1.7	40.6	36	24.7	18	9.0	8.4
1969	532.6	767.7	14.1	18.8	6.6	6.0	2.7	1.7	42.7	32	29.7	18	9.4	9.5

* Divisor is negative or negligible.

Subtracting the estimate of direct taxes attributable to labor income from total direct taxes leaves direct taxes attributable to property income. A precise allocation of direct personal taxes to the net return from corporate, noncorporate, and other assets is impossible to achieve. As a plausible first approximation, we allocate the remainder of direct taxes proportionately to the net return from noncorporate capital, the net return to corporate capital after corporate profit taxes, and income from claims on governments and foreigners (approximated by net interest paid by federal, state, and local governments). The estimate of direct personal taxes, allocated to all property income is given in the eleventh column of Table 1. An effective rate of direct personal taxation on property income can be estimated by dividing the taxes in column (11) by the appropriate base. The base includes the net corpo-

---

1951–58 figures show an average proportion of .755 with a negligible amount of variation. Since the relative proportions of labor and capital income were stable for the same period and have continued to remain stable, it is reasonable to assume that the .755 proportion held from 1959–69.

rate return less property taxes and corporate income taxes, the net noncorporate return less property taxes, and net interest paid by federal, state, and local governments. The estimated effective rate is given in the twelfth column of Table 1.

We can now proceed to compute after-tax rates of return. The noncorporate after-tax return is the net return less property taxes and a proportionate share of direct personal taxes. Dividing this after-tax return by the value of noncorporate assets at the end of the previous year yields the after-tax rate of return shown in column (13). The corporate after-tax return is the net return less property taxes, corporate income taxes, and a proportionate share of direct personal taxes. Dividing the after-tax return by the value of corporate assets yields the after-tax rate of return shown in column (14).

Comparing the estimated after-tax rates of return for the forty-one years in our sample, we find that the noncorporate rate of return exceeds the corporate rate in twenty-one of the years, is exceeded by the corporate rate in eighteen of the years, and is equal to the corporate rate in 1952 and 1967. The arithmetic means of the series are extremely close, 7.6 for the noncorporate sector and 7.3 for the corporate sector. The correlation coefficient between the series is .93. The period from 1929 through the early postwar years can hardly be classified as economically stable; for information on equilibrium rates of return most observers would give more weight to the postwar period. It is precisely the period from 1947 to 1969 that seems to best substantiate our hypothesis of rates of return tending to equality.

IV. Conclusion and Remarks

Under the assumptions made in the previous sections the conclusion is unmistakable that entrepreneurial income measures up very well. This is contrary to all previous discussions of this aspect of social accounting for the United States.

This conclusion evokes two obvious questions: What accounts for the disparity between this finding and those of previous studies? What biases may be involved in the calculations leading to this conclusion? In answering these questions it is convenient to compare the methods used here with those of Kravis.

Since our labor imputation is exactly the same as that used by Kravis, the discrepancy between our conclusions must lie in the treatment of property income. Kravis' methodology appears to be essentially the same as ours. After imputing labor income to entrepreneurs, Kravis imputes income to entrepreneurial capital and finds that total entrepreneurial income is not large enough to fulfill both imputations. This implies that we should find the noncorporate rate of return is dominated by the corporate rate of return—but we do not. Thus we must examine more closely Kravis' imputation to entrepreneurial capital. We quote the full description from Kravis: "The entrepreneurial property share in national income can then be calculated by applying to the share of other property (i.e., the combined share of interest, rent and corporate profits) the ratio of the value of entrepreneurial to other property" (p. 924). The main problem is immediately clear. Since national income is the basis for computation, the imputation is gross of all direct taxes and net of capital gains. Thus Kravis is implicitly assuming either that rates of return before taxation and exclusive of capital gains tend to equality, or that rates of capital gains and direct taxation are the same in both sectors. The former assumption is at odds with received economic theory, which we have discussed above; and the latter assumption is without empirical foundation. There may be other shortcomings in the imputation, but it is difficult to tell from the terse description provided. For example, if the "other property" share of national income includes government interest in addition to corporate interest, then the "other" rate of return is overstated because there is no corresponding property.

Regardless of any other possible discrepancies in the capital imputation, not accounting for differences in taxation and capital gains goes a long way to explaining the difference between our results and those of Kravis. Imputing income to entrepreneurial capital using the corporate rate of return before the corporate income tax involves a serious error.[7]

If the corporate and noncorporate capital stocks had roughly the same mix of structures, equipment, inventories, and land, the inclusion or exclusion of capital gains would make little difference. However, the asset mixes are not similar; an important difference is that land accounts for approximately 50 percent of the noncorporate asset value but only 15 percent of the corporate asset value. The price of land has behaved quite differently from the prices of the other assets. In particular it has increased more rapidly in the postwar period.[8] For the entire sample period the average annual capital gain was 3.6 percent for the noncorporate sector and 2.9 percent for the corporate sector. Thus noncorporate capital did not have to generate as large a current income flow to provide a rate of return comparable to that of corporate capital.

The reason for the discrepancy with the results of Kravis (and others who used his methodology) is clear. Kravis has erred in his attempt to impute a competitive rate of return to entrepreneurial capital. This error has caused the illusion that entrepreneurial income has been insufficient to provide competitive returns for both labor and capital input. The conclusion argued here is the opposite. We have found no evidence that entrepreneurial income has been insufficient to provide competitive returns to entrepreneurial labor and capital.

The conclusion of this study is, of course, conditional on the conventions and assumptions used to construct the basic data. We have already mentioned some possible biases and we now point out some others.

It might be argued that the effective rate of personal income taxation on corporate net income is lower than on noncorporate capital income. The chief argument for this view would be that corporate retained earnings produce capital gains to corporate shares, which are not taxed until realized. Furthermore, long-term capital gains are taxed at preferential rates and intergenerational transfer of corporate shares permit complete avoidance of the personal income tax. If this and other methods of avoiding and delaying personal taxes on corporate net income is superior to methods for delaying and avoiding taxes on noncorporate income, then the proportional allocation of personal taxes tends to overstate the noncorporate rate of return and understate the corporate rate. This is an important question, but its detailed examination is beyond the scope of this paper. One piece of evidence suggests, however, that this is not the source of a large bias. In the postwar period dividends paid have averaged 36 percent of the net return to corporate capital.[9] This large portion is subject to the personal income tax. In view of the large capital gains on noncorporate assets, it would be surprising if the portion of the net return to noncorporate capital paid out as ordinary income greatly exceeded the 36 percent of corporate net return paid as dividends. Even if some reallocation of taxes were necessary, it would not affect our conclusion unless it were quite large.

---

[7] Kravis is certainly not the only writer who has committed this error.

[8] See Christensen and Jorgenson (Tables 2 and 3) for figures supporting these assertions.

[9] Note that "net return" refers to the numerator used in computing the rate of return in the seventh column of Table 1—not to corporate profits.

A bias may exist due to errors in measurement of asset values used in the rate of return calculations. There may be errors in the procedures used by Christensen and Jorgenson in measuring the capital stocks, but two exceptions aside, we have not been able to identify any systematic errors in one direction. One exception is the inclusion of the entire agricultural capital stock in the noncorporate sector, while part of the corresponding income flow is allocated to the corporate sector.[10] This biases the estimated corporate rate of return upward and the noncorporate rate of return downward; the error is quite small early in the period but becomes larger in recent years with the rapid growth of corporate farming.[11]

The second bias is similar—the neglect of existing capital stock which passes from one sector to the other. This is not accounted for because we have measured capital stocks by cumulating depreciated investment. Existing unincorporated business firms often incorporate, but the converse rarely occurs. The resulting bias again tends to overstate the noncorporate capital stock and understate the corporate capital stock. Hence the noncorporate rate of return is underestimated and the corporate rate of return is overestimated. Corrections for both of these biases would tend to increase the noncorporate rate of return and decrease the noncorporate rate.

Finally, let us suppose that our conclusion is correct, i.e., that refinements of the data led to an even closer agreement between the corporate and noncorporate rates of return. What implications would this have for economic arguments regarding the motivations of self-employed workers? One argument says that the expected rate of return to noncorporate capital must exceed that of corporate capital because noncorporate investments entail more risk. Another argument says, "The advantages of being one's own master surely exist and are surely positive. If they operated as other equalizing advantages do, they should work to keep down the dollar returns from self-employment" (Lebergott, p. 82). Our conclusion does not imply that these arguments are invalid. It does imply that if they do exist, they offset each other.

REFERENCES

A. Ando and E. C. Brown, "Lags in Fiscal Policy," in E. C. Brown et al., eds., *Stabilization Policies, Commission on Money and Credit,* Englewood Cliffs 1963.

L. R. Christensen and D. W. Jorgenson, "The Measurement of U.S. Real Capital Input, 1929–1967," *Rev. Income & Wealth,* Dec. 1969, *15,* 293–320.

E. F. Denison, "The Sources of Economic Growth in the United States and the Alternatives Before Us," Supplementary Paper No. 13, Committee for Economic Development, New York 1962.

———, *Why Growth Rates Differ: Postwar Experience in Nine Western Countries,* Washington 1967.

L. Frane and L. R. Klein, "The Estimation of Disposable Income by Distributive Shares," *Rev. Econ. Statist.,* Nov. 1953, *35,* 333–37.

---

[10] The allocation of investment by sector is based on Goldsmith, who treats agriculture as a third sector distinct from the corporate and noncorporate sectors. The allocation of all agricultural investment to the noncorporate sector was the best approximation that could be made.

[11] The growing importance of corporate farming can be seen by comparing corporate sales in agriculture, forestry, and fisheries with the total value of farm output. From 1929–50 corporate sales averaged approximately 5 percent of the total value of farm output. During the 1950's this percentage gradually increaased, and exceeded 10 percent for the first time in 1960. In 1967, the 20 percent figure was reached. See Tables 1.17 and 6.19 in *OBE* (1966).

R. W. Goldsmith, *The National Wealth of the United States in the Postwar Period,* Nat. Bur. Econ. Res., *Stud. in Capital Formation and Financing,* Princeton 1962.

D. G. Johnson, "The Functional Distribution of Income in the United States, 1850–1952," *Rev. Econ. Statist.,* May 1954, *36,* 175–82.

I. B. Kravis, "Relative Income Shares in Fact and Theory," *Amer. Econ. Rev.,* Dec. 1959, *49,* 917–49.

S. Kuznets, "Quantitative Aspects of the Economic Growth of Nations, IV," *Econ. Develop. Cult. Change,* Apr. 1959, *7,* 1–100.

S. Lebergott, "Factor Shares in the Long Term: Some Theoretical and Statistical Aspects," in *The Behavior of Income Shares: Selected Theoretical and Empirical Issues,* Nat. Bur. Econ. Res., *Stud. in Income and Wealth,* Vol. 27, Princeton 1964.

A. D. Manvel, "Trends in the Value of Real Estate and Land, 1956–1966," in *Three Land Research Studies,* The National Commission on Urban Problems, Res. report no. 12, 1968.

L. G. Rosenberg, "Taxation of Income from Capital, by Industry Groups," A. C. Harberger and M. J. Bailey, eds., *The Taxation of Income from Capital,* Washington 1969, 123–84.

R. C. Wasson, J. C. Musgrave, C. Harkins, "Alternative Estimates of Fixed Business Capital in the United States, 1925–1968," *Surv. Curr. Bus.,* Apr. 1970, *50,* 18–36.

U.S. Bureau of Labor Statistics, *Handbook of Labor Statistics,* Washington 1968.

U.S. Office of Business Economics, *National Income 1954, Surv. Curr. Bus., supp.,* Washington 1955.

———, *The National Income and Product Accounts of the United States, 1929–65, Surv. Curr. Bus., supp.,* Washington 1966.

Substitution, Complementarity, and the Residual Variation Around Dynamic Demand Equations

By Louis Phlips*

This paper reports on efforts made to analyze the residual variation around demand equations, obtained by maximizing a directly additive and "dynamic" quadratic utility function. We start from a model constructed by H. S. Houthakker and L. D. Taylor (1966, 1970, ch. 5), which introduces this utility function. We show that, under certain conditions, the sign of the correlations between the residuals obtained after jointly estimating all the corresponding demand equations provides valuable information on the sign of the (total) cross substitution effects and, consequently, on the substitution and complementarity relationships among pairs of commodities. We also argue that these signs indicate the presence or absence of misspecification with regard to want independence. A multistage principal component analysis of the residual correlation matrix is then carried out, in a search for the level of aggregation best suited for the use of directly additive utility functions.

On the basis of the two-stage maximization hypothesis, we finally analyze the residual variation in subsets of demand equations, to measure substitution and complementarity relationship within more disaggregated commodity sets.

I. The Houthakker-Taylor Model

Houthakker and Taylor (hereafter HT) consider a consumer who can purchase n commodities and whose preferences may be described by a utility function of the form

$$\phi(q, s) = q'a + s'b + \tfrac{1}{2}q'Aq + q'Bs + \tfrac{1}{2}s'Cs \tag{1}$$

where q and s are $(n\text{x}1)$ vectors of purchases and "state variables," respectively, a and b are constant vectors, and A, B, and C are constant nonsingular and diagonal matrices; q and s are functions of time, which is treated as continuous.

The state variables $s = (s_1, \ldots, s_n)'$ are interpreted as in HT (1970, pp. 9–10) and stand for nonmeasurable stocks of durable goods or habits. The vector Bs measures the influence of habit formation (or inventory adjustment) on marginal utilities: the model thus introduces a deterministic explanation of the changes in preferences over time. This explanation seems very plausible and is certainly an improvement in comparison to the introduction of a time trend.

HT assume the state variables to be related to the flows through

$$\frac{ds}{dt} = q - Ds \tag{2}$$

where D is assumed to be diagonal and

---

* Professor of economics, Catholic University of Louvain. I am most grateful to H. S. Houthakker and L. D. Taylor for constant guidance and encouragement. Thanks are due to A. P. Barten, J. Drèze, J. Garcia dos Santos, and J. Waelbroeck for helpful comments on earlier drafts, and to Denzo Kamiya, Ann Walka, J. Selvidge, and Betty Carrol for research and programming assistance. The computations were carried out during my stay at Harvard University as a Fulbright visiting scholar and were financed by a National Science Foundation grant to Harvard University. All errors are mine.

nonsingular. The third and final equation in the model is the budget constraint

$$p'q = y \tag{3}$$

HT then maximize (1) subject to (3). Differentiating the result with respect to time and using (2) to eliminate the unobservable s, they obtain after some manipulation

$$\frac{dq}{dt} = -A^{-1}Da - (A^{-1}B + D)q + A^{-1}\frac{dp^*}{dt} + A^{-1}Dp^* \tag{4}$$

where $p^* = \lambda p$. The marginal utility of income, λ, appears as a linear function of income y and of s.

For empirical purposes, a finite approximation of the preceding results is obtained with the help of the procedure outlined in HT (1970, pp. 13–16). The estimating equation turns out to be

$$q_t = K_0 + K_1 q_{t-1} + K_2 \lambda_t p_t + K_3 \lambda_{t-1} p_{t-1} \tag{5}$$

where

$$K_0 = -[I + (1/2)(A^{-1}B + D)]^{-1}A^{-1}Da \tag{6}$$

$$K_1 = [I + (1/2)(A^{-1}B + D)]^{-1} \cdot [I - (1/2)(A^{-1}B + D)] \tag{7}$$

$$K_2 = [I + (1/2)(A^{-1}B + D)]^{-1} \cdot A^{-1}(I + (1/2)D) \tag{8}$$

$$K_3 = -[I + (1/2)(A^{-1}B + D)]^{-1} \cdot A^{-1}(I - (1/2)D) \tag{9}$$

Given the diagonality of A, B, and D, (5) is seen as the equation for the ith good, all matrices being treated as scalars, and equations (6) to (9) can be solved for the structural coefficients a, A, B, and D.

Differentiation of the model leads to short- and long-run price and income derivatives.

In the short run (defined as a period such that $\partial s/\partial y = \partial s/\partial p = 0$), HT obtain, from

$$a + Aq + Bs = \lambda p \tag{10}$$

and from (3), the following results

$$\lambda_y = \frac{\partial \lambda}{\partial y} = (p'A^{-1}p)^{-1}, \tag{11}$$

a negative scalar, as A is negative definite,

$$q_y = \frac{\partial q}{\partial y} = (p'A^{-1}p)^{-1}A^{-1}p \tag{12}$$

and

$$q_p = \frac{\partial q}{\partial p'} = \lambda A^{-1} - \lambda(p'A^{-1}p)^{-1}A^{-1}pp'A^{-1} - \frac{\partial q}{\partial y}q' \tag{13}$$

To obtain long-run derivatives, use is made of equation (2). One obtains

$$\lambda^*y = \frac{\partial \lambda^*}{\partial y} = (p'W^{-1}p)^{-1}, \tag{14}$$

where $W = A + BD^{-1}$ and stars indicate long-run equilibrium values,

$$\overset{*}{q}_y = \frac{\partial q^*}{\partial y} = (p'W^{-1}p)^{-1}W^{-1}p \tag{15}$$

and

$$\overset{*}{q}_p = \frac{\partial q^*}{\partial p'} = \lambda W^{-1} - \lambda(p'W^{-1}p)^{-1}W^{-1}pp'W^{-1} - \frac{\partial q^*}{\partial y}q' \tag{16}$$

To estimate the coefficients K_0, K_1, K_2, and K_3, (5) is rewritten as

$$q_{it} = K_{i0} + K_{i1}q_{i(t-1)} + K_{i2}\lambda_t p_{it} + K_{i3}\lambda_{t-1}p_{i(t-1)} + \epsilon_{it}, \tag{17}$$

there being one equation for each commodity i. These n equations are estimated jointly, with the help of an iteration on λ_t insuring that the adding-up criterion is satisfied. Imposing $\sum \hat{q}_{it} = y_t$, (where q and

$$\lambda_t = \frac{y_t - \sum \hat{K}_{i0} - \sum \hat{K}_{i1} q_{i(t-1)} - \lambda_{t-1} \sum \hat{K}_{i3} p_{i(t-1)}}{\sum \hat{K}_{i2} p_{it}} \tag{18}$$

---

y are measured as expenditures in constant prices per capita) an estimate of λ can indeed be obtained from (18), where the "hats" denote estimates. At the first step of the iteration, λ_t is set equal to 1 for all t in (17). The estimated K's and the data are used to estimate a first set of λ's according to (18) and this set is used to adjust the prices. Equations (17) are re-estimated using the adjusted prices, and so on, until the process converges.

II. A Stochastic Model

In the HT model, changes in preferences are explained by the deterministic influence of s on the marginal utilities. As for the disturbances ϵ, the usual assumptions of classical least squares are supposed to be satisfied, possible autocorrelation being eliminated or at least reduced by the "three-pass least squares" method of Lester Taylor and Thomas Wilson.

It is tempting to go one step further and to view these disturbances as changes in the purchases resulting from random fluctuations in the marginal utilities. The idea is to locate these random shocks in the linear part a of the utility function, the matrices A and B being considered as constant over time, and to admit that *the off-diagonal elements in A can be non-zero,* in order to stay within the general case.

We accordingly admit that

$$\epsilon_t = \Delta q_t = \frac{\partial q}{\partial a'} \Delta a'_t \tag{19}$$

and want to determine $\partial q/\partial a'$ on the assumption that the consumer continues to act rationally, i.e., to maximize utility within the budget constraint. Our reasoning is thus entirely in line with the approach outlined by Henri Theil and H. Neudecker for the static quadratic utility function and further developed by A. P. Barten (ch. 3) and Theil (1967, ch. 7).

Starting from the first-order conditions (3) and

$$a + Aq + Bs = \lambda p, \tag{20}$$

we differentiate with respect to a, treating p and y as constant and λ and q as variable.

In the short run, $\partial s/\partial a = 0$. Differentiation of (20) and (3) gives

$$I + A \frac{\partial q}{\partial a'} = p \frac{\partial \lambda}{\partial a'} \tag{21}$$

and

$$p' \frac{\partial q}{\partial a'} = 0 \tag{22}$$

This leads to

$$\frac{\partial q}{\partial a'} = A^{-1}\left(p \frac{\partial \lambda}{\partial a'} - I\right) \tag{23}$$

$$p' \frac{\partial q}{\partial a'} = p'A^{-1}p \frac{\partial \lambda}{\partial a'} - p'A^{-1} = 0 \tag{24}$$

$$\frac{\partial \lambda}{\partial a'} = (p'A^{-1}p)^{-1}p'A^{-1} \tag{25}$$

and

$$\frac{\partial q}{\partial a'} = -[A^{-1} - (p'A^{-1}p)^{-1}A^{-1}pp'A^{-1}] \tag{26}$$

which, when compared with (13), is seen to be the short-run total price substitution effect except for sign and for λ (which is positive). Equation (26) may be rewritten as

$$\frac{\partial q}{\partial a'} = -[A^{-1} - (1/\lambda_y)q_y q_y'] \tag{27}$$

In the long run, we have

$$I + A\frac{\partial q^*}{\partial a'} + B\frac{\partial s^*}{\partial a'} = p\frac{\partial \lambda^*}{\partial a'} \tag{28}$$

$$p'\frac{\partial q^*}{\partial a'} = 0 \tag{29}$$

and

$$\frac{\partial s^*}{\partial a'} = D^{-1}\frac{\partial q^*}{\partial a'} \tag{30}$$

which implies

$$\frac{\partial q^*}{\partial a'} = (A + BD^{-1})^{-1}\left(p\frac{\partial \lambda^*}{\partial a'} - I\right) = W^{-1}\left(p\frac{\partial \lambda^*}{\partial a'} - I\right) \tag{31}$$

$$\frac{\partial \lambda^*}{\partial a'} = (p'W^{-1}p)^{-1}p'W^{-1} \tag{32}$$

$$\frac{\partial q^*}{\partial a'} = -[W^{-1} - (p'W^{-1}p)^{-1}W^{-1}pp'W^{-1}] = -[W^{-1} - (1/\overset{*}{\lambda}_y)\overset{*}{q}_y\overset{*}{q}_y{}'] \tag{33}$$

In both the long and the short run, the random shocks influence the quantities demanded in the same way as the substitution effect of price changes, except for sign and a positive constant. We conclude that the error term of (17), which is short-run, is

$$\epsilon_{it} = -\sum_j\left[a^{ij} - (1/\lambda_y)\frac{\partial q_i}{\partial y}\frac{\partial q_j}{\partial y}\right]\Delta a_{jt} \tag{34}$$

where a^{ij} indicates an element of A^{-1}.

We now introduce a probability distribution for the random shocks and for the ϵ's, at each t and for each j, and proceed to derive the moments. Following Barten (par. 3.3), we suppose $E(\Delta a_{jt}) = 0$ such that

$$E(\epsilon_{it}) = 0, \tag{35}$$

and also that the "contemporaneous" variance-covariance matrix of the random shocks can be specified for any t, as

$$E[(\Delta a_t)(\Delta a_t)'] = -k_t^2 A \tag{36}$$

where k is a non-zero scalar. As A is negative definite by assumption, the minus sign is needed to insure positive variances. The specification says that the variances are proportional to the absolute value of the second-order direct partial derivatives of the utility function: if the marginal utility of good i is much affected by a change in its quantity, then the variance of Δa_{it} is large. As for the covariances, these are zero in case of want independence. Otherwise, they are proportional to the cross-partial derivatives of the utility function, a negative derivative implying a positive correlation.

It follows, then, that the contemporaneous covariance matrix of the ϵ's is

$$E(\epsilon_t\epsilon_t') = -k_t^2[A^{-1} - (1/\lambda_y)q_yq_y']A \cdot [A^{-1} - (1/\lambda_y)q_yq_y']$$

as $[A^{-1} - (1/\lambda_y)q_yq_y']$ is symmetric, or

$$E(\epsilon_t\epsilon_t') = -k_t^2[A^{-1} - (1/\lambda_y)q_yq_y'] \cdot [I - (1/\lambda_y)Aq_yq_y'] \tag{37}$$

However,

$$(1/\lambda_y)Aq_y = (p'A^{-1}p)A(p'A^{-1}p)^{-1}A^{-1}p = p \tag{38}$$

Therefore

$$E(\epsilon_t\epsilon_t') = -k_t^2[A^{-1} - (1/\lambda_y)q_yq_y'] \cdot [I - pq_y'] \tag{39}$$

But, as can be verified,

$$[A^{-1} - (1/\lambda_y)q_yq_y'][-pq_y'] = 0 \tag{40}$$

Finally,

$$E(\epsilon_t\epsilon_t') = -k_t^2[A^{-1} - (1/\lambda_y)q_yq_y'] \tag{41}$$

The contemporaneous covariances of the

error terms are seen to be equal to the total cross substitution effects except for sign and a positive constant. When two goods are substitutes in the Hicksian sense, the total cross substitution effect is positive and the correlation of the corresponding error terms is negative. When two goods are complements in the Hicksian sense, the correlation is positive. And when *all* goods are independent

$$\text{cov}(\epsilon_{it}, \epsilon_{jt}) = (k_t^2/\lambda_y) \frac{\partial q_i}{\partial y} \frac{\partial q_j}{\partial y} \tag{42}$$

which are all negative (as $\partial\lambda/\partial y$ is negative) in the absence of inferior goods. Except for sign and a positive constant, these are what Houthakker called "general" substitution effects in his 1960 article.

These findings imply that, if good estimates of the ϵ's can be obtained, the (contemporaneous) correlation matrix of these estimates gives valuable information on the structure of preferences. If all correlations turn out to be negative, we are not allowed to reject the hypothesis that A is diagonal. If the signs differ, we may interpret a negative correlation between ϵ_{it} and ϵ_{jt} to indicate that goods i and j are substitutes, and a positive that they are complements. But then we have to admit a misspecification of the utility function.

III. Estimation Procedures

Our procedure to estimate the correlation matrix of the ϵ's will be simplistic: we will admit that the correlations of the residuals obtained after estimating (17) by three-pass least squares, at the last step of the iteration on λ, are good estimates of the correlations of the ϵ's. To justify this, the following arguments might be advanced.

First of all, it might be recalled that systematic changes in preferences are taken into account by the model, through the influence of s. If the residuals were nevertheless autocorrelated, this autocorrelation would have been eliminated by three-pass least squares.

Second, there is the possibility that a specification error is made when A is supposed to be diagonal, i.e., when equation (5) is treated as the equation for the ith good and written as (17). This error does not affect the estimates of the residuals if the omitted variables influence the dependent variables only through the included independent variables. However, if the omitted variables also have a direct impact on the dependent variables, then residuals on the average will be too large. The sign of the residual correlations may nevertheless remain unaffected. Our naive procedure has then the merit of exploiting the advantages (from the point of view of estimation) of the additivity assumption without loss of information as to possible substitution or complementary relationships.

Third, it is clear that the true disturbances of the estimating equations may contain components other than the influence of the random shocks as defined in (34). One might think of a long list of errors, such as (further) specification errors, aggregation errors, observational errors, etc. But one might argue, also, that these errors are not contemporaneously correlated.

Finally, it might be that the preceding arguments would be convincing only if the results obtained seem plausible. They do seem plausible, however.

Once the residual correlation matrix (R) is available, a natural proceeding is to factorize it into principal components,[1] to get a clearer picture of the possible groupings of the commodities. As principal component analysis is becoming more and more familiar, it is sufficient to recall the essential steps of the computations (for further details, see Richard Stone).

---

[1] Principal component analysis is preferred to (common) factor analysis because the former analyzes the total variation and does not imply any a priori grouping.

Consider a set of standardized (last step) residuals $x_{it}(i=1, \ldots, n;\ \mathrm{t}=1, \ldots, T)$. We wish to compute unobserved orthogonal factors (principal components) best capable of linearly "explaining" the variation of the x_i. It is desired that each x_i shall be as nearly as possible reproduced by $f_{i1}z_1$, where the coefficients f_{i1} are constants and z_1 is the first principal component (varying through time). This amounts to finding n coefficients f_{i1} and T values $z_{1\mathrm{t}}$ such that

$$\sum_i \sum_{\mathrm{t}} (x_{it} - f_{i1}z_{1\mathrm{t}})^2 \tag{43}$$

is a minimum. Solving the minimizing equations one finds

$$z_{1\mathrm{t}} = \frac{\sum_i f_{i1}x_{it}}{\sum_i (f_{i1})^2} \tag{44}$$

where $\lambda_{\mathrm{t}}^* = \sum_i (f_{i1})^2$ is the largest latent root of the correlation matrix of the x's. The vector f_1 is the first of n orthogonal vectors to which correspond n (orthogonal) principal components. The second principal component corresponds to the second highest latent root λ_2^* of the correlation matrix, etc. As the sum of the latent roots equals the trace of R, the principal components exhaust the data. For standardized x's the coefficients f are simple correlation coefficients between the x's and the z's. Following standard psychometric terminology we shall call them "factor loadings." The ratio of each λ_j^* to n measures the proportion of the total variation (the trace of R) accounted for by the corresponding principal factor. If there is a grouping of the commodities, this is revealed by the fact that they have high correlations (f_{ij}) with the same factor z_j.

We might have stopped the analysis here, the factorization of the residual correlation matrix giving a clear picture of the structure of preferences within groups of commodities. However, we thought it might be interesting to try to determine the relationships, if any, between the groups (i.e., principal components). Suppose all components are negatively correlated with each other: this might indicate that the corresponding commodity groups are independent. The argument is intuitive, though, and should perhaps not be taken too seriously.[2]

As principal components are orthogonal by construction, they have to be rotated in some way (to make them oblique) if one wants to analyze the correlation between them. Looking into a huge psychometric literature (see, e.g., H. H. Harman), we found that there are very many ways indeed to rotate components. It was decided to use the so-called Biquartimin version of J. B. Carroll's Oblimin program (see Harman, ch. 15), which had the advantage of being available and recommended by most factor analysts. Call F_1 the matrix of factor loadings expressing the simple correlation between the x's and their principal components. This program rotates F_1 and sets up the matrix of intercorrelations Φ_1 among the resulting oblique factors. Needless to say that Φ_1 can be factorized again into principal components, which can be rotated again, etc. The procedure is continued until Φ becomes a unit matrix.[3]

---

[2] If $z_{\mathrm{t}}^* = G\epsilon_{\mathrm{t}}$ is a vector of rotated components (at time t), $E(z_{\mathrm{t}}^* z_{\mathrm{t}}^{*\prime}) = -k_{\mathrm{t}}^2 G[A^{-1} - (1/\lambda_v)q_v q_v']G'$. The correlations between components are thus seen to be weighted sums of all the (own and cross) substitution effects.

[3] It might be interesting to notice that these are the essential steps involved in the calculation of the "hierarchical factor solution" which is well known to psychometrists (see J. Schmid and J. M. Leiman). This solution gives an orthogonal picture of the successive higher-order factors, using a theorem on orthogonalization by L. R. Tucker. However, it has the disadvantage of analyzing only the common factor space.

Table 1—Correlation Matrix (R), Eleven Variables (Last-Step Residuals)

	1	2	3	5	4	8	6	7	9	10
1. Auto & Parts	1.0000									
2. Furniture Equipment	−0.5920	1.0000								
3. Other Durables	−0.8558	0.5606	1.0000							
5. Clothing & Shoes	−0.8163	0.3570	0.7203	1.0000						
4. Food & Beverage	−0.3065	−0.2784	0.3172	0.3120	1.0000					
8. Housing	−0.1329	0.3915	0.0292	0.2371	−0.4273	1.0000				
6. Gasoline & Oil	0.1844	−0.4630	−0.2185	0.0419	0.2495	−0.0302	1.0000			
7. Other Nondurables	−0.2201	0.1731	0.0129	0.0309	0.0772	−0.0436	−0.1645	1.0000		
9. Household Operation	0.1040	0.0908	−0.0098	−0.1238	−0.2643	0.1052	0.2012	−0.0051	1.0000	
10. Transportation	−0.0573	−0.1064	0.1550	−0.0401	0.1933	−0.2660	0.0422	−0.0960	0.1611	1.0000
11. Other Services	0.0950	−0.3358	−0.1329	−0.1553	−0.1865	−0.3779	−0.0518	−0.2634	−0.3030	−0.2171

IV. Empirical Results

The United States

The procedure outlined above was first applied to last step residuals obtained after fitting (17) to the eleven consumption series (in real terms per capita) published in the *Survey of Current Business* (for the years 1947–64) and iterating on λ, the price for each commodity being the ratio of the implicit deflator for the commodity to the deflator for total consumption expenditure.

As for the regression results, it may suffice here to report the excellent fitting obtained: all coefficients were highly significant, the R^2's ranged from .81 to .99 and autocorrelation was eliminated by the use of three-pass least squares.

The analysis of the residual correlation matrix gives encouraging results (Table 1).

First of all, we notice that the correlations between commodities are not all negative, as would have been expected with an additive preference ordering. Almost one half (32 out of 55) are positive.

Furthermore, looking at the matrix of factor loadings (Table 2) we find an interesting and clear-cut grouping. Four durable commodities (cars, furniture and household equipment, other durables, clothing and shoes) are highly correlated with the first factor. It is worth noting that clothing, which is classified in the *Survey of Current Business* under nondurable goods, belongs here to the group of durables. The second factor corresponds to food and housing, the third picks up three services while the fourth and fifth correspond to

Table 2—First-Order Factor Loadings (F_1), Eleven Variables, *U.S.* Consumption Data

	1	2	3	4	5	6	7	8
Cumulative Percent of Trace	29.65	48.39	61.59	73.05	83.31	89.18	93.91	97.20
1. Auto & Parts	0.9237	0.1972	0.1263	0.0746	0.0022	0.0428	0.1591	0.0213
2. Furniture & Household Equipment	−0.7416	0.4966	−0.0673	−0.1196	0.0093	0.2191	0.0729	0.3388
3. Other Durables	−0.8860	−0.2216	−0.0214	−0.1691	−0.2101	−0.0510	0.0171	0.1098
4. Food & Beverage	−0.2244	−0.8178	0.2334	0.0817	0.1966	−0.1221	0.3627	−0.0771
5. Clothing & Shoes	−0.8181	−0.2578	−0.1616	0.2498	−0.1290	−0.1823	−0.1226	−0.1807
6. Gasoline & Oil	0.2835	−0.4619	0.0699	0.7057	−0.0774	−0.0592	−0.3137	0.3033
7. Other Nondurables	−0.2217	0.0981	0.1868	−0.1179	−0.8864	−0.0268	−0.3121	−0.0568
8. Housing	−0.2909	0.6371	−0.0991	0.5991	−0.1198	0.1141	−0.0141	−0.2815
9. Household Operation	−0.0003	0.4680	0.5819	−0.1691	−0.2508	−0.5809	−0.0929	0.0263
10. Transportation	−0.0400	−0.3006	0.6790	−0.2507	−0.3474	0.4220	−0.2588	−0.1316
11. Other Services	0.3473	−0.2310	−0.7079	−0.4248	−0.1947	−0.1122	−0.2665	−0.0664

gasoline and oil and other nondurables, respectively.

What are the relationships *within* these five groups? After rearranging the correlation matrix we find indication (Table 1)—looking at the signs—that cars (1) are substitutes for all other durables (2, 3, 5) whereas furniture (2) is complementary to "other durables" (3) and clothing (5). "Food and beverages" (4) is negatively correlated with housing (8). In the last two columns, transportation (10) appears as a complement to household operation (9) and as a substitute to "other services" (11), with which household operation is also negatively correlated. All this seems quite plausible.

As for the relationships between items belonging to different groups, we pinpoint complementarity between furniture (2) and housing (8), substitutability between furniture and household equipment (2) and gasoline and oil (6), and substitutability between housing services (8) and other services (11).

These results seem to indicate that even the five groups are not independent. To obtain better information on this point, the five first factors of Table 2 were rotated. The matrix of intercorrelations (Φ_1) among the rotated factors is shown in Table 3 and is close to the unit matrix. Most of these correlations are nonnegative.

Factorizing Φ_1 gives the principal components in Table 4. The first takes the nondurables (food, housing, other) together while the three following factors correspond each to a first-order component. An oblique rotation[4] leads to Φ_2 (Table 5) which is a unit matrix. It is impossible to rotate any further: the variables are orthogonal. If anything, this implies that even a grouping of consumer expenditures into four groups is not likely to satisfy the additivity assumption.

TABLE 3—MATRIX OF INTERCORRELATIONS (Φ_1) AMONG FIRST-ORDER ROTATED FACTORS

	1	2	3	4	5
1. Durables	1.0000	0.0088	0.0673	0.0659	0.1421
2. Food & Housing	0.0088	1.0000	0.0699	−0.1886	0.1542
3. Services	0.0673	0.0699	1.0000	−0.0129	0.0650
4. Gas & Oil	0.0659	−0.1886	−0.0129	1.0000	0.0153
5. Other Nondurables	0.1421	0.1542	0.0650	0.0153	1.0000

TABLE 4—SECOND-ORDER FACTOR LOADINGS (F_2)

	1	2	3	4	5
Cumulative Percent of Trace	25.68	48.95	68.00	85.11	100.00
1. Durables	−0.3455	−0.6311	−0.1327	−0.6692	−0.1303
2. Food & Housing	−0.6907	0.3664	−0.1386	0.1405	−0.5914
3. Services	−0.4106	−0.1836	0.8856	0.0968	0.0630
4. Gas & Oil	0.3635	−0.6938	−0.0091	0.4528	−0.4258
5. Other Nondurables	−0.6221	−0.3406	−0.3623	0.4164	0.4386

TABLE 5—MATRIX OF INTERCORRELATIONS (Φ_2) AMONG SECOND-ORDER ROTATED FACTORS

	1	2	3	4
1. Nondurables	1.0000	−0.0471	0.0848	−0.0436
2. Gas & Oil	−0.0471	1.0000	−0.0024	0.0644
3. Services	0.0848	−0.0024	1.0000	−0.0024
4. Durables	−0.0436	0.0644	−0.0024	1.0000

To test this interpretation, the eleven expenditures series were combined into four groups corresponding to our four (third-order) components. The correlation matrix of the residuals obtained after jointly estimating the dynamic demand equations (17) was factorized into principal components. It was hoped that each item would have corresponded to a different component, but these hopes were not realized. However, of the six correlations, only four were negative, which makes it difficult again to accept the hypothesis of additive preferences.

---

[4] It is readily admitted that an arbitrary element creeps in at the moment where only five out of the eleven columns of F_1, and four out of the five columns of F_2 are rotated. The application of a test of significance would have introduced some statistical rigor. It was decided not to apply such tests as they imply large samples and as their use in factor analysis is questionable anyway. Common sense seems to be the only criteria available in the present state of the art. And on this basis one could argue that the percent of variation left out is precisely the irrelevant part of the total variance.

The Durable Goods Group

The analysis of substitution and complementarity relationships is not too fascinating when commodities are highly aggregated. But it becomes particularly interesting when specific types of goods can be recognized. We decided therefore to take a closer look at some of the eighty-four consumption series out of which the *Survey of Current Business* constructs the eleven series used above.

It is an impossible task to jointly estimate eighty-four demand equations. Is one allowed to analyze a certain number of commodities and how does one proceed? Some theorizing is required at this point.

A useful theoretical framework is provided by Robert Strotz's two-stage maximization hypothesis, according to which households proceed in two steps, the first being an allocation of income among broad commodity groups corresponding to branches of the utility function, the second step implying the optimal spending of each budget allotment in its branch, with no further reference to purchases in other branches.[5]

In his 1957 article, Strotz defined the utility functions as being weakly separable. William Gorman tried to show—and Strotz (1959) subsequently agreed with this—that the two-stage procedure has to be based, for three or more branches, on a strongly separable, i.e., additive, utility function. Recently, however, Charles Blackorby, George Lady, David Nissen and R. Robert Russell made it clear that the Strotz-Gorman conditions refer only to the existence of group price indices and showed that the budgeting procedure described by Strotz is consistent if, and only if, the utility function is weakly separable.

The empirical results of the previous subsection are in keeping with this new formulation of the hypothesis, as no evidence pointing towards strong separability could be found. We felt justified, therefore, to proceed on the assumption that households allocate income among broad groups and maximize a quadratic "branch" utility function as defined in (1), given a budget constraint equal to total expenditures in a particular group, where q, s, and p are now $(m\text{x}1)$ vectors, m being the number of items in the group.

On a first trial, we analyzed some of the groups as given in the *Survey*, in particular the groups "transportation," "recreation," "financial services." The results were completely implausible for those groups which included a mixture of durables, nondurables and services.

We then decided to take the results of the multistage factor analysis more seriously and to admit an allocation of income among the final four groups. The analysis of the group "durables," including twelve items, is reported here and is very encouraging indeed.[6]

The residual correlation matrix (Table 6) indicates the following complementary pairs (limiting ourself to correlations higher than .6): clothing (2) and jewelry (3), clothing (2) and semi-durable house furnishing (8), jewelry (3) and furniture (4), jewelry (3) and toys, sports equipment, boats, etc. (11). As substitutes we have: clothing (2) and cars (9), jewelry (3) and cars (9), household appliances (5) and cars (9), semi-durable house furnishing (7) and toys, sports equipment, boats, etc. (11).

A principal component analysis (Table 7) isolates a first subbranch including

---

[5] For a review of the literature on this subject, see H. A. John Green.

[6] All slope coefficients are significant and have the right sign, the R^2's ranging from .79 to .97.

TABLE 6—CORRELATION MATRIX (R), BRANCH "DURABLE GOODS" (1947–64)[a]

	2	3	4	5	8	9	6	7	11	1	10	12
2.	1.0000											
3.	0.6023	1.0000										
4.	0.4087	0.6581	1.0000									
5.	0.4163	0.3737	0.3798	1.0000								
8.	0.7847	0.3865	0.3086	0.3004	1.0000							
9.	−0.7127	−0.6251	−0.5611	−0.6402	−0.7901	1.0000						
6.	−0.0532	−0.0428	0.2096	0.5983	0.1461	−0.4333	1.0000					
7.	−0.1570	−0.1065	0.2323	0.2906	0.1467	−0.3548	0.5414	1.0000				
11.	0.5500	0.6620	0.3469	0.2527	0.2130	−0.2737	−0.1362	−0.6761	1.0000			
1.	−0.3247	0.0153	−0.1423	−0.0662	−0.3636	0.0745	−0.0313	0.0369	−0.2472	1.0000		
10.	0.0721	0.2160	0.1966	−0.0344	0.0344	−0.3336	0.2429	−0.0246	0.2789	−0.2231	1.0000	
12.	0.1716	0.1225	−0.0291	0.3559	0.1789	−0.4050	0.0407	0.4213	−0.1652	−0.0500	−0.0035	1.0000

[a] Categories are identified in Table 7.

clothing, jewelry, furniture, household appliances, semi-durable house furnishing and cars. A second group includes china, other durable house furnishing and toys, sports equipment, etc. Shoes, tires and auto parts form a third group, while radio and TV receivers stay apart. As seen in Tables 8 and 9, higher order F matrices lead to a distinction between what may be called "basic" durables and luxuries.

Food in the United Kingdom

The *U.S.* series do not give any detail on food expenditures. We therefore turned to the ten *Blue Book* series for the United Kingdom, imposing a budget constraint equal to total expenditure on food.[7] With

TABLE 7—FIRST-ORDER FACTOR LOADINGS (F_1) OBTAINED BY FACTORIZING THE BRANCH "DURABLE GOODS" (1947–64)

	1	2	3	4	5
Cumulative Percent of Trace	35.43	55.82	68.24	77.04	84.40
1. Shoes & other footwear	0.2556	−0.1514	−0.6558	−0.4765	−0.1534
2. Clothing, including luggage	−0.8221	0.3117	0.2911	−0.1108	0.1165
3. Jewelry and watches	−0.7618	0.3375	−0.2351	−0.2006	−0.2638
4. Furniture	−0.6823	0.0218	−0.2491	0.2060	−0.1926
5. Kitchen and other HH appliances	−0.6777	−0.3549	−0.0367	0.1672	−0.4539
6. China, glassware, tableware & utensils	−0.3337	−0.6514	−0.2912	0.5186	0.0566
7. Other durable house furnishing	−0.1606	−0.9159	0.0343	−0.0180	0.0403
8. Semi-durable house furnishing	−0.7482	−0.0073	0.4012	−0.0324	0.4087
9. New cars & net purchases of used cars	0.9355	0.2407	0.0522	0.1382	−0.1566
10. Tires, tubes, accessories & parts	−0.2578	0.0494	−0.7102	−0.1092	0.5469
11. Wheel goods, durable toys, sports equipment boats & pleasure aircr.	−0.5263	0.7455	−0.1719	0.1266	−0.1376
12. Radio & TV receivers; records & musical instruments	−0.2997	−0.4552	0.2700	−0.6235	−0.1376

---

[7] The slope coefficients are not significantly different from zero for bread and fish. The R^2's are the following: .17 for fish, .41 for oils, .65 for sugar, and higher than .84 for all other goods.

TABLE 8—SECOND-ORDER FACTOR LOADINGS (F_2)
BRANCH DURABLE GOODS

	1	2	3	4
Cumulative Percent of Trace	32.77	60.05	83.92	100.00
1. Clothing, furniture, cars, etc.	−0.3360	−0.6934	−0.5587	−0.3069
2. China, other house furnishing, sports equipment, etc.	−0.7028	−0.3968	0.3527	0.4735
3. Shoes & tires	0.4233	−0.5534	0.6546	−0.2933
4. Radio & TV records & musical instruments	−0.7243	0.3733	0.2995	−0.4886

TABLE 9—THIRD-ORDER FACTOR LOADINGS (F_3)
BRANCH DURABLE GOODS

Cumulative Percent of Trace	38.97	72.97	100.00
1. China, etc. & radio, etc.	0.0846	0.9552	0.2837
2. Clothing, furniture, etc.	−0.7511	0.2811	−0.5974
3. Shoes & tires	−0.7733	−0.1685	0.6113

only nine observations and a budget constraint limited to what is but a sub-group of the branch "nondurables," the grouping should not be expected to be too clear-cut.

The reader may judge the results (Tables 10 and 11) for himself. We find them surprisingly good: meat and fish come out as substitutes, fish and oils as complements, fish and fruit as unrelated, bread and potatoes as substitutes, etc. High elasticity goods stand together against basic food. Finally, this distinction into high and low elasticity goods is confirmed by the second-order factor loadings (Table 11).

V. Conclusions

Starting from Houthakker and Taylor's system of dynamic demand equations, derived from a generalized additive quadratic utility function, we have shown that under certain not too restrictive assumptions the residual correlations give valuable information on the (total) cross substitution effects. In particular, the signs of the correlations turn out to be the opposite of those of the substitution effects.

The residual correlations of a complete system of eleven demand equations for the United States are such that the hypothesis of additive preferences cannot be accepted at this level of aggregation. A higher-order factor analysis seems to suggest that this hypothesis remains unrealistic at higher levels of aggregation as well.

Plausible substitution and complementarity relationships between the eleven aggregates are detected. On the basis of the two-stage maximization hypothesis, an attempt is also made to measure these relationships between more disaggregated commodities within the group "durables"

TABLE 10—FIRST-ORDER FACTOR LOADINGS (F_1), 10 VARIABLES,
FOOD IN THE UNITED KINGDOM

	1	2	3	4	5
Cumulative Percent of Trace	31.77	60.50	76.52	85.76	93.26
1. Bread	−0.3993	−0.6247	−0.3147	−0.5674	−0.0933
2. Meat	−0.9077	0.0038	0.2031	−0.0912	0.0070
3. Fish	0.7716	0.4148	−0.0384	−0.4121	0.9950
4. Oils	0.6393	−0.2922	−0.4277	−0.0481	0.4692
5. Sugar & preserves	−0.0197	−0.8353	0.3583	0.1029	0.2704
6. Dairy	−0.5075	0.2929	0.6054	0.0725	0.3436
7. Fruit	0.5589	−0.5228	0.1706	0.3203	−0.4980
8. Potatoes	−0.0068	0.8149	−0.3866	0.3805	0.0886
9. Beverages	0.2775	0.6850	0.4896	−0.3971	−0.2064
10. Other	−0.7336	0.1907	−0.6033	−0.0161	−0.1476

TABLE 11—SECOND-ORDER FACTOR LOADINGS (F_2), FOOD IN THE UNITED KINGDOM

	1	2	3
Cumulative Percent of Trace	38.61	75.13	100.00
Meat, fish, etc.	0.8425	0.0633	0.5350
Bread, sugar, etc.	0.5350	−0.6904	−0.4869
Dairy	−0.4030	−0.7841	0.4719

(for the United States) and within the group "food" for the United Kingdom. The empirical results are surprisingly good.

Given the (apparently insurmountable) difficulty of estimating complete systems of demand equations without the assumption of additive preferences, and the resulting loss of information on cross-relationships,[8] it is concluded that the analysis of the residual variation around demand equations derived from an additive quadratic utility function provides an interesting and rewarding approach to the empirical analysis of substitution and complementarity relationships.

REFERENCES

A. P. Barten, "Theorie en empirie van een volledig stelsel van vraagvergelijkingen," doctoral dissertation, Rotterdam 1966.

C. Blackorby, G. Lady, D. Nissen and R. R. Russell, "Homothetic Separability and Consumer Budgeting," *Econometrica,* May 1970, *38,* 468–72.

W. M. Gorman, "Separable Utility and Aggregation," *Econometrica,* July 1959, *27,* 469–81.

H. A. J. Green, *Aggregation in Economic Analysis, An Introductory Survey,* Princeton 1964, ch. 2.

H. H. Harman, *Modern Factor Analysis,* Chicago 1960.

H. S. Houthakker, "Additive Preferences," *Econometrica,* Apr. 1960, *28,* 244–57.

——— and L. D. Taylor, *Consumer Demand in the United States, 1929–1970,* 2d ed., Cambridge, Mass. 1970.

——— and ———, "Joint Estimation of Dynamic Demand Equations," mimeo., 1966.

J. Schmid and J. M. Leiman, "The Development of Hierarchical Factor Solutions," *Psychometrika,* 1957, *22,* 53–61.

J. R. N. Stone, "On the Interdependence of Blocks of Transaction," *J. Roy. Statist. Soc.,* 1947 supp., *9,* pt. 1, 1–32.

R. H. Strotz, "The Empirical Implications of a Utility Tree," *Econometrica,* 1957, *25,* 269–80.

———, "The Utility Tree—A Correction and Further Appraisal," *Econometrica,* July 1959, *27,* 482–88.

L. D. Taylor and T. A. Wilson, "Three-Pass Least Squares: A Method for Estimating Models with a Lagged Dependent Variable," *Rev. Econ. Statist.,* Nov. 1964, *46,* 329–46.

H. H. Theil, *Economics and Information Theory,* Amsterdam 1967.

——— and H. Neudecker, "Substitution, Complementarity, and the Residual Variation around Engel Curves," *Rev. Econ. Stud.,* Feb. 1958, *25,* 114–23.

L. R. Tucker, "The Role of Correlated Factors in Factor Analysis," *Psychometrika,* 1940, *5,* 141–52.

U.S. Office of Business Economics, *Surv. Curr. Bus.,* 1947–64.

United Kingdom, Central Statistical Office, *National Income and Expenditure* (Blue Book), London 1955–63.

---

[8] As is well known, additive preferences imply that all goods appear as Hicks-Allen substitutes.

Debt in the Consumer Portfolio: Evidence from a Panel Study

By WILLIAM C. DUNKELBERG AND FRANK P. STAFFORD*

Consumer behavior is frequently described in the context of a general portfolio model which specifies equilibrium levels of consumer assets and liabilities. The principal arguments in the demand functions for various consumer assets are long-run factors such as normal income and its growth over time and a vector of yields (or relative prices) on consumer assets. For a given set of equilibrium conditions, the concern then is with the rate of adjustment toward equilibrium for any observed disequilibrium in the portfolio. Such stock adjustment models, however, are rarely subjected to empirical testing at the micro-economic level due to the lack of conformable data (in this case, a complete reinterview panel measuring "stocks" of assets and liabilities at two points in time).

This paper presents the empirical results of a study of the use of consumer debt in the framework of a stock adjustment model of behavior. Data are from a two-year reinterview panel (begun in 1967) financed by the Ford Foundation and conducted by the Survey Research Center, University of Michigan, under the supervision of the authors and other staff members at the Center.[1] The model permits estimates of the direct impact of the stock of debt on net changes in instalment liabilities (new borrowing less repayments) as well as the indirect impacts of disequilibria in stocks of liquid assets and consumer durable goods (the value of cars and household durables owned). The effects of the maturity composition of instalment obligations and the labor force participation of other family members relative to expected levels are also included in the analysis. The model is based on the findings presented by Harold Watts and James Tobin and entails a "... 'general disequilibrium' framework for the dynamics of adjustment to a 'general equilibrium' system" (p. 106).

I. The Model

The general stock adjustment model can be specified as follows:

$$E_i = a_{i0} + \sum_{j=1}^{n} a_{ij}(S_j^* - S_j) + e_i$$

where

E_i = increment to the ith stock
a_{ij} = direct ($i=j$) and cross ($i \neq j$) adjustment coefficients
S_j^* = the desired value of the jth stock
S_j = the actual value of the jth stock

The complete model contains as many equations as stocks (n).

The relationship to be estimated in this paper is as follows:

---

* Assistant professor of business economics, Graduate School of Business, Stanford University and assistant professor of economics, University of Michigan, respectively. We would like to thank members of the Microeconomic Research and Public Policy Seminar at Stanford, the Money Seminar at Michigan, and the referee for helpful comments. An earlier version of this paper was presented at the Econometric Society Meetings in December 1969.

[1] The panel was a national reinterview sample of 1,921 *U.S.* households *excluding* households with family heads over sixty years of age. Whenever applicable, actual dollar amounts were ascertained and not interval bracket values.

$$\begin{aligned}\Delta D = {} & a_0 + a_1(D^* - D_{-1}) \\ & + a_2(H^* - H_{-1}) \\ & + a_3(A^* - A_{-1}) \\ & + a_4\left(\left(\frac{P}{Y}\right)^* - \left(\frac{P}{Y}\right)_{-1}\right) \\ & + a_5(W^* - W_{-1}) + \mu \end{aligned} \tag{1}$$

where

ΔD = net change in instalment credit during the year (extensions − repayments)

D^* = predicted equilibrium level of instalment debt

H^* = predicted equilibrium level of household durables (including automobiles)

A^* = predicted equilibrium level of liquid assets defined to include savings bonds, checking accounts, and savings accounts

$(P/Y)^*$ = predicted equilibrium level of the ratio of annual debt payments to income

W^* = predicted level of wife's (and other's) labor earnings

The subscript -1 denotes initial values of the variables.

Generally speaking, the use of instalment credit provides a means for the consumer to adjust expenditures so that consumption is smoothed over time and does not become dependent on the more uneven flows of resources (see Irving Fisher, James Tobin, Milton Friedman, Watts and Tobin). Borrowing from income flows over the near term is particularly common for younger families who are in the process of accumulating an inventory of household durables. These families will use credit even at a relatively high cost because the expected rate of return from investing in certain household durables is very high (see Stafford and Dunkelberg for a discussion of costs of credit).[2]

In addition to borrowing for household durables, consumers can use credit to pay for training. Various studies have shown that increments to human wealth through training are widespread phenomena and that younger families are particularly likely to engage in training and incur debt in the process (see Gary Becker, John Johnstone and Ramon Rivera, Jacob Mincer (1962)). The borrowing for training need not be direct since financing the purchase of durables will free cash for such alternative uses.

Although families do borrow to help achieve their particular consumption (and investment) objectives, the borrowing process is not the sole means for doing so, and, in part, equilibrium is achieved through short-run adjustments in the work-leisure balance of the household. Arnold Zellner has argued that ". . . the actual leisure-work balance may be different from that desired by the household and this would prompt a change in the extent of labor force participation. Even more important, however, in influencing the labor force participation may be deviations of actual stocks from desired levels, particularly as regards durable goods and debt" (1960, p. 489, fn. 5). This kind of short-run adjustment in the wife's participation would be an alternative to increased borrowing, particularly if the family has already gone into debt and is facing high marginal borrowing costs. In the literature on secondary earners (see Glen Cain and Mincer (1960)), the desire of the family to stabilize consumption is a motivating factor which operates to increase the short-run supply of labor by the wife, although local labor market demand conditions may have the opposite effect in that the demand for the husband and wife's services are similarly affected over the cycle.

---

[2] The annual rate of return to owners of an automatic washer and dryer has been estimated to range from 1 to 27 percent, depending primarily on the number of children in the family. See James Poapst and William Waters.

In a study of the permanent income hypothesis by Robert Holbrook and Stafford, analysis of panel data revealed that there is in fact a negative correlation between "unexpected" short-run changes in the husband's earnings and short-run changes in the wife's (and others') earnings. Moreover, the association between short-run changes in the husband's and wife's earnings was somewhat greater for families with heads under 45 years of age than for families with heads over 45.[3] This stronger correlation can serve as a measure of the extent to which variations in the labor supply of secondary earners are used as an alternative to variations in borrowing and in liquid assets. For younger families with small children who have a more limited financial portfolio, adjustments in the short-run work-leisure balance, particularly by the wife, are a more important means of stabilizing family welfare over time, even though the presence of small children operates to limit this means of adjustment.

The fact that consumers hold both debt and liquid assets when they could presumably retire their debt at the expense of reduced liquid assets is a well recognized phenomenon. This implies that consumers do view the existence of such assets as a means of producing utility, a form of consumption. This choice is described in more general terms as "liquidity preference," or the "demand for money." As Tobin has argued,

> There are minimal requirements of gross liquid asset holdings. . . . These requirements may make it worthwhile to incur the cost of simultaneous holding of liquid assets and debt . . . since the household's income could stand as a measure of the scale of transactions and contingencies for which liquidity is required, this optimum could be expressed as a proportion of income. . . . Households who have to borrow to obtain liquidity would settle for less gross liquidity than wealthier spending units. In this context, credit allows for the payment of unexpected and/or indivisible expenditures to be spread out over a time, easing the impact on the existing stock of liquid assets. [p. 533]

Based on these general considerations, the desired level of earnings of the wife, W^*, and the optimum ratio of annual debt payment commitments to disposable income, $(P/Y)^*$, were included with the more conventionally used stocks of durables, H^*, debt, D^*, and liquid assets, A^*. House value and mortgage debt were excluded from the analysis since such stocks have a much longer adjustment period and therefore are less a part of the shorter run portfolio adjustment process.

Operationally, the desired or equilibrium level of each stock is its expected value as determined by its empirical relationship to a set of exogenous variables. Underlying this approach is the assumption that the panel members are on average in equilibrium at a particular point in time, and that properly conditioned means yield estimates of the equilibrium values. Since there are lags in the adjustment processes, it is more likely that consumers will be systematically above or below equilibrium levels. For example, in periods of rapidly rising interest rates, observed consumer stocks of debt may be too high relative to equilibrium levels (since the process of debt reduction takes time). Using observed levels of debt to estimate desired or equilibrium levels will likely lead to systematic overstatements of the desired portfolio levels and an understatement of the disequilibrium.

---

[3] The correlations of short-run changes in husband's and wife's income components by various age and education subgroups are not reported in the paper and are as follows (1959–1960): under 45, less than high school education, −.43; under 45, high school education or greater, −.17; over 45, less than high school education, −.09; over 45, over high school education or more, −.14.

An alternative approach which avoids this difficulty is to assume a particular form of the adjustment process (for example, a first-order difference equation with proportional adjustment), estimate the adjustment parameters based on observations on the change in each stock and then infer the equilibrium values from the solution of the equations. This approach has the disadvantage of restricting the nature of adjustment process. One objective of our work was to test for nonlinearities in the adjustment process (there may be limitations on the rate of adjustment of extreme disequilibria). A priori specification of the exact form of the adjustment process precludes this possibility and focuses on the equilibrium values themselves, rather than the adjustment process.

The equilibrium values of the stocks were specified as follows:

$$D^* = d(Y^*\ X_1\ X_2\ X_3) \tag{2}$$

$$H^* = h(Y^*\ X_3\ X_4\ X_5) \tag{3}$$

$$A^* = a(Y^*\ X_1\ X_2\ X_3) \tag{4}$$

$$\left(\frac{P}{Y}\right)^* = p(Y^*\ X_1\ X_2\ X_3) \tag{5}$$

$$W^* = w(X_6\ X_7\ X_8\ X_9\ X_{10}\ X_{11}) \tag{6}$$

Normal income was estimated as follows:

$$Y^* = y(X_{12}\ X_{13}\ X_{14}\ X_{15}) \tag{7}$$

where

D^* = the desired level of instalment debt
H^* = the desired value of the stock of durables
A^* = the desired level of liquid assets (savings accounts, bonds, and checking accounts)
$(P/Y)^*$ = the optimum size of total monthly debt payments relative to disposable income
W^* = the desired level of the wife's earnings
Y^* = normal income
X_1 = 1 if a farmer, 0 otherwise
X_2 = 1 if the respondent reports large monthly income variations, 0 otherwise
X_3 = family life cycle (see detail in Table 1)
X_4 = 1 if owns a home, 0 otherwise
X_5 = 1 if rents *furnished*, 0 otherwise
X_6 = age of the wife
X_7 = education of the wife
X_8 = city size of residence
X_9 = age of the youngest child
X_{10} = number of children
X_{11} = hourly wage of the head
X_{12} = age of the family head
X_{13} = education of the family head
X_{14} = occupation of the family head
X_{15} = house value *or* monthly rent (see Tables 1 and 2 for a description of the forms of these variables)

Equations (2) to (6) were estimated by regressing observed values of the stock variables at the beginning of the period on the initial values of the predictors (the results are presented in Tables 1 and 2). Residuals, then measure the disequilibria specified in the central equation (1). These disequilibria were then used to predict net changes in outstanding credit obligations during the next twelve months (ΔD).

Normal income, Y^*, is the prime determinant of all equilibrium values of stocks, and relative prices are assumed to be the same for all consumers.[4] The arguments for such a specification are spelled out by

---

[4] As Mincer has pointed out (1963), this is not strictly correct. Although it is reasonable to assume that *market* prices do not vary systematically in cross sections, *full* prices of goods, meaning market prices plus opportunity costs of time used in consuming goods will vary systematically by, say, income class. For our purpose biased income elasticities are less important than the need for income to predict desired levels of stocks.

Zellner. Normal income functions were estimated separately for home owners and for renters by regressing total family income on a conventional set of characteristic variables and on house value or monthly rent, respectively. The values predicted by the equation were then transformed to interval values to allow for nonlinearities (see Table 1 for a description of the form of normal income).

The desired level of instalment debt, D^*, and the optimum annual payment-income ratio, $(P/Y)^*$, in equations (2) and (5) depend on longer run income, variations in income, and on family life cycle (a composite of age, marital status, age and presence of children and labor force status). The level of instalment debt and the size of monthly payments are different aspects of a family's commitment to time purchases. There will be an increasing burden on liquidity as the level of payments relative to income increases (see John Lansing, E. Scott Maynes, and Mordechai Kreinin). Thus, in any one year, even though extensions of credit may be relatively low, extensions in subsequent periods may be lower than expected if the maturity composition of the current stock of debt is relatively short (high monthly payments). The payments mechanism also increases the sensitivity of spending to actual or expected income changes. Expenditures can be increased by a multiple of the increase in debt payments that new income will support.[5]

The expected size of the stock of liquid assets, A^*, as specified by equation (4) depends primarily on normal income, a measure of the scale of transactions and contingencies for which such assets are required. Families whose incomes vary from month to month would desire larger equilibrium levels of liquid assets.[6] The size of this stock will vary over the family life cycle as needs for liquidity, income, and the composition of total assets (liquid and non-liquid) change.

Household durables stocks, H^*, were defined to include the market value of automobiles, televisions, refrigerators, stoves, and washing machines owned. The expected value of this stock was specified to depend on normal income, home ownership, family life cycle, and for renters, whether or not the dwelling unit is furnished. Renters of furnished dwelling units would be expected to have smaller desired stocks since they are contracting for a major part of their durables consumption and this is reflected in their rent payments.

A household's short-run adjustment response also depends on the actual and desired work-leisure balance of the household. This study focuses on the wife's labor force participation. The expected level of earnings from her employment (equation (6)) is dependent on her inventory of human capital (approximated by her age and education), labor market opportunities (approximated by the degree of urbanization of her area of residence), her alternative productivity in the household (age and number of children), and the earning potential of her husband (hourly wage). Since a substantial amount of the variation in the husband's earnings is due to variations in hours, his hourly wage was used as a measure of earning potential.

II. Empirical Results

Equilibrium Levels of Selected "Stocks"

The estimation of the desired levels of the stocks described in equations (2)–(6) is shown in Tables 1 and 2. Dummy variable coefficients were estimated under the constraint that their sum, weighted by the size of each class, equals zero for each predictor. This yields the mean value of the dependent variable as the regression constant term and provides a coefficient for each class of each predictor. The few

---

[5] The expansion of consumer demand in 1955 is attributed in part to changes in credit terms.

[6] For an "inventory" type model as it applies to firms, see Merton Miller and Daniel Orr.

TABLE 1—PREDICTED LEVELS OF HOUSEHOLD INVENTORIES AND THE DEBT PAYMENT TO INCOME RATIO

Independent Variables	PCT. DIST.	Dependent Variables			
		(2) Instalment Debt	(3) Autos & Other Consumer Durables	(4) Liquid Assets	(5) Debt Payment to Income Ratio
Constant Term		$813 (1238)	$2071 (1351)	$2544 (7852)	8 percent (10 percent)
Normal Income[a]					
Under $4,000	8.4	−54 (93)	−753 (81)	−2318 (581)	3.3 (.7)
$4,000–6,999	25.2	−75 (54)	−517 (47)	−1567 (335)	1.9 (.4)
$7,000–9,999	36.3	76 (45)	31 (39)	−535 (280)	1.0 (.3)
$10,000–14,999	20.8	30 (59)	392 (51)	2144 (369)	−2.0 (.5)
$15,000 or more	9.3	−110 (95)	1094 (83)	3650 (595)	−4.1 (.7)
Family Age and Composition[b]					
Under Age 45					
Unmarried, no children	4.8	−289 (123)	−346 (169)	374 (768)	−1.5 (.9)
Married, no children	6.8	100 (104)	388 (90)	504 (647)	0.7 (.8)
Married, youngest child under 6	28.5	217 (51)	−16 (44)	−794 (316)	2.1 (.4)
Married, youngest child over 6	14.4	179 (71)	116 (61)	−794 (444)	0.9 (.5)
Age 45 or Older					
Married, children present	17.5	78 (65)	96 (56)	−205 (402)	−0.2 (.5)
Married, no children, head in labor force	14.4	−208 (71)	254 (62)	1226 (444)	−2.1 (.5)
Married, no children, head retired	1.0	−252 (275)	−399 (238)	4393 (1715)	−0.4 (2.1)
Unmarried, no children, head in labor force	5.4	−586 (116)	−663 (100)	2033 (723)	−4.2 (.9)
Unmarried, no children, head retired	1.0	−665 (282)	−848 (244)	298 (1755)	−0.9 (2.1)
Any Age					
Unmarried, has children	6.3	−337 (109)	−429 (107)	−153 (676)	−0.8 (.8)
Whether Farmer					
Farmer	4.0	−562 (136)		1401 (85)	−5.1 (1.0)
Not a farmer	96.0	23 (27)		−58 (17)	0.2 (.2)
Whether Income Varies from Month to Month					
Varies	18.7	60 (63)		−267 (390)	0.7 (.5)
Does not vary	81.3	−14 (30)		61 (187)	−0.2 (.2)
Housing Status					
Renter	29.3		−23 (67)		
Owner	70.7		10 (28)		
Whether Apartment Furnished					
Rents furnished	3.9		−234 (118)		
Owns or unfurnished	96.1		10 (24)		
R (adjusted)		.211	.490	.247	.274

[a] F test for this set of variables significant at 1 percent level in equations (3), (4), (5).
[b] F test for this set of variables significant at 1 percent level in equations (2), (3), (4), (5).
Note: Standard errors in parentheses.

variables that could be represented in continuous form (such as normal income) were used in categorical form for ease of estimation and to allow for non-linear relationships.

The desired level of instalment credit bears a non-linear relationship to normal or permanent income as estimated by equation (2). In addition to the effect of income, the presence of young children was associated with increased levels of instalment credit. Farm families were found to have lower levels of instalment credit as expected. Presumably this lower level for farmers is related to the irregular pattern of farm income during the year and to the resulting difficulty in meeting regular payments. In an attempt to extend the measurement of income variability, respondents were asked if their income varied from month to month. Income variation as measured by the response to this question had only a minor relation to predicted levels of instalment credit.

The effect of normal income is the most pronounced for predicting the household's durable inventory. This finding is consistent with previous work on durables (see Gregory Chow) in which normal income is used to explain the consumption of durable goods.[7] In addition to normal income, a family's age and structure, X_3, has a bearing on desired durable inventories. The impact of these variables on long-run income is included in Y^* but the desired path of the level of durables consumption (and all consumption for that matter) varies as a function of family structure quite apart from normal income (see Martin David). This is because both the permanent income and life cycle hypotheses of consumption do not account for expected changes in family size over time, or changes in the composition of consumption which alter spending patterns. Even though a single individual may have a low current income relative to his income expected in the near future, he also expects to have increased family obligations which will tend to reduce *per capita* income. In this regard it is not surprising that younger unmarried individuals, desiring to smooth per capita consumption over time, consume less household durables at a given level of normal income (see Stafford).

In previous research there has often been an owner-renter distinction and we have followed this precedent only insofar as the owner-renter distinction has an informational value in assigning normal income. Once desired levels of the consumer portfolio were specified, we did not hypothesize any great difference in the overall adjustment process at least with regard to instalment credit.

The level of liquid assets, defined as the sum of checking and savings accounts and savings bonds, was most highly dependent on normal income. Taking the midpoint of the \$4,000–6,999 normal income bracket as \$5,500 with its corresponding predicted liquid asset level of \$977 (\$2,544–1,567) and the midpoint of the \$10,000–14,999 normal income bracket as \$12,500 with its corresponding predicted liquid asset level of \$4,688 (\$2,544 +2,144), the arc income elasticity of liquid assets over this range can be calculated. The resulting estimate is about 1.7. While this estimate is rather imprecise, it indicates that the income elasticity of liquid asset holdings is quite high as measured by our regression coefficients.

In addition to income effects, one can also observe the effects of family composition on liquid assets. The pattern of liquid asset change can be summarized as follows: liquid assets are built up by single

---

[7] Taking the midpoint of the \$4,000–6,999 bracket as \$5,500 and taking the midpoint of the \$10,000–14,999 bracket as \$12,500 and taking predicted inventory values of \$1,550 and \$2,460, we have an approximate arc elasticity over this range of 1.7 which compares favorably with the range estimates by Chow of 1.4 to 2.0 for automobiles.

TABLE 2—PREDICTED LEVELS OF THE WIFE'S EARNINGS (6)

Independent Variables	PCT. DIST.	Coefficient (Std. Error)
Age of Wife		
Under 25	9.5	$258 (151)
25–34	28.9	281 (87)
35–44	29.5	170 (86)
45–54	25.2	−350 (93)
55 or older	6.9	−970 (177)
Education[a]		
0–5 grades	2.5	−910 (308)
6–8 grades	10.8	−422 (142)
9–11 grades	20.1	−430 (104)
12 grades	34.3	−159 (80)
High School plus training	13.5	382 (128)
College, no degree	10.8	123 (142)
College, B.A.	7.2	1345 (174)
College, advanced degree	0.7	6660 (560)
City Size		
Central cities, 12 largest	9.5	169 (152)
Cities 50,000 or over except 12 largest	18.6	145 (109)
Urban places 10,000–49,999	18.6	150 (108)
Urban places 2,500–9,999	23.4	−12 (97)
Rural, in an SMSA area	5.8	47 (194)
Rural, not in an SMSA area	23.7	−319 (96)
Not ascertained	0.4	1207 (770)
Number of Children Under 18		
None	26.4	337 (91)
One	23.8	64 (96)
Two	21.8	−123 (100)
Three	14.2	−247 (124)
Four	7.0	−335 (178)
Five or more	6.9	−272 (179)
Age of Youngest Child		
Less than 2 years	12.5	−816 (132)
2–2.9 years	6.4	−549 (185)
3–3.9	5.6	−562 (197)
4–4.9	6.0	−505 (191)
5–5.9	5.3	−316 (203)
6–8.9	11.9	−144 (136)
9–13.9	16.7	21 (114)
14–18	9.2	717 (157)
No children under 18	26.4	637 (91)
Hourly Wage of the Head		
Did not work	13.1	−248 (129)
$0.50–0.99	3.0	251 (268)
$1.00–1.49	4.2	222 (215)
$1.50–1.99	6.7	238 (181)
$2.00–2.49	8.0	516 (165)
$2.50–2.99	10.5	382 (144)
$3.00–3.49	14.3	−49 (124)
$3.50–4.49	19.8	−14 (105)
$4.50–5.99	12.8	−186 (131)
$6.00–9.99	7.6	−526 (170)

Constant term = $1422 *R* (adjusted) = .417

[a] *F*-test for this set of variables significant at the 1 percent level.
Note: Standard errors in parentheses.

individuals and young marrieds with no children and decrease as young children appear and consumption needs rise, and then increase once again as the children grow older and prior to retirement. The effect of monthly income variation is opposite to what one might expect from inventory type models of the demand for money. However, it is of very small magnitude (statistically insignificant) and another measure of income variability, the farm-nonfarm occupation variable does have the expected sign and is of substantial magnitude.[8]

The equilibrium debt payment-to-income ratios are inversely related to normal income, suggesting that the average level of monthly payments rises less than proportionately as one goes up the income scale. In all other respects, payment to income ratios vary in the same way as does the level of instalment credit.

The predictive equation for the wife's earnings (Table 2) correspondents to a well-developed literature on the labor force participation of the wife. Basically, her potential market productivity as measured by her education and location of residence is positively related to her earnings while the value of her alternative home produc-

---

[8] One problem with farm assets is the extent to which they finance business rather than personal transactions.

TABLE 3—STOCK ADJUSTMENT OF INSTALMENT CREDIT[a]

	(D^*-D_{-1})	(H^*-H_{-1})	(A^*-A_{-1})	$\left(\left(\frac{P}{Y}\right)^*-\left(\frac{P}{Y}\right)_{-1}\right)$	(W^*-W_{-1})	$(\bar{R}^2)$
All Panel Families	0.473 (.028)	0.040 (.023)	0.008 (.003)	4.500 (3.318)	−0.040 (.014)	.238
Debtors Only	0.517 (.034)	0.036 (.034)	.015 (.009)	15.018 (4.036)	− .043 (.020)	.284
Debtors Only	—	0.350 (.034)	—	—	− .037 (.022)	.07

[a] Standard errors in parentheses.

tion as measured by the number of children and age of the youngest child is negatively related to her earnings. The husband's hourly wage is inversely related to the wife's earnings. One interesting result is the consistent decline in the wife's earnings with age. It is possible that the woman's withdrawal from the labor force reduces her market productivity since she fails to engage in the training normally a part of most jobs.[9]

Stock Adjustment Estimates of Instalment Credit Changes

Given the predicted values of instalment credit, household physical and financial assets, payment to income ratios, and the wife's earnings from equations (2) to (6), the adjustment equation, (1), was estimated. The net change in instalment credit outstanding between early 1967 and early 1968 was regressed on deviations between expected and actual values of the stock variables as of January 1967.[10] The results are summarized in Table 3 with the estimates for the entire sample of 1,921 families in row 1 and the results for only the 1,377 families with debt in either year given in row 2.

Here the estimated adjustment coefficient of instalment credit for a given gap between desired and actual levels is about .5 whether we use all families in the panel (equation (1)) or only the debtors (equation (2)). The interpretation is that half of the difference between actual and desired levels of instalment credit is reduced in a given year. This result can be compared with estimates from other studies. For example, Zellner, in summarizing the Watts and Tobin paper at the 1960 Wharton School Conference commented that "... the instalment debt process seems, ..., to be independent of everything but the auto stock adjustment process. These findings are hard to believe ... " (p. 498).

Here we find that for debtors the adjustment process is not only highly dependent on the disequilibrium with respect to instalment credit itself, but also depends on disequilibrium in the payment to income ratio and in the wife's earnings. However, the other cross-adjustment effects, particularly those for household durables and assets, have much less impact than does the direct adjustment effect. One plausible reason for these smaller effects of household durable and liquid assets and the puzzling results in the Watts and Tobin study is the fact that a disequilibrium in debt serves as a good measure for corresponding disequilibria in both of these other sectors of the portfolio. A family with unusually low current debt could be, at the same time, above its usual (or desired) level of liquid assets and also below its usual (or desired) level of household durables. If in the subsequent period new durables are bought, debt will increase and some liquid assets will also be used up to help finance the purchase. As a result, desired debt being considerably above actual debt may serve as a proxy for an abnormally low level of household inventories which should induce acquisition of new goods in the next period and the incurrence of new debt (here it is assumed that liquid assets and debt are used in combination to finance expenditures). If our measures of desired levels of household durables (and liquid assets) are measured less completely than instalment credit, then correlations between the true disequilibria and that of instalment credit can confound the measured cross-adjustment coefficients.

---

[9] For a theoretical model relating human capital to life cycle increments to the human capital stock, see Yoram Ben Porath. This model suggests an exogenously given rate of depreciation of human capital. Hence, women out of the labor force would experience a continuous decline in human capital, with no offsetting increase in human capital through on-the-job training. Our empirical result applies to a cross section. The general growth of the economy should operate to offset this decline with age by shifting the earnings profile upward over time (see Becker (1964)) as would a secular shift in the labor force participation of women. The general level of female labor force participation can be analyzed with a general model of the allocation of time within the household (see Becker (1965)).

[10] For the wife's earnings the relevant data would be for the previous year, 1966.

If instalment credit disequilibrium is the mirror image of the household durable disequilibrium then it is in the same sense redundant to use both to predict changes in debt. To determine the effect of household durable disequilibria on instalment credit extensions, we estimated our adjustment equation with only household durables and wife's earnings disequilibria as independent variables. The results, summarized by equation (3) of Table 3 do indicate that there is a direct relationship between adjustments in the stock of durable goods and net credit use.

The adjustment effects which are unique to this study are those for the payment to income ratio and the wife's earnings. In Table 3 it can be seen that families with high initial payment to income ratios reduced their credit obligations, and conversely families with low payment to income ratios increased their credit obligations. This finding would imply that if per capita instalment credit had remained stable but credit terms were tightened and only shorter maturities were available, there would be a decline in credit extensions during the subsequent period(s).

Variation of the wife's labor force participation (or more generally, labor supply) is an important part of the process of the stabilization of the family's resources over time. We find that there is a trade off between changes in credit obligations and changes in the wife's labor supply. If the wife is at "full employment," that is, working as much as or more than "normal," the family will increase its debt obligations to finance expenditures and/or liquidity during the period. In contrast, if the wife is employed at a level below her expected level, the desire to finance new expenditures and/or liquidity will tend to encourage her to increase her labor supply. From this analysis one can see a fairly direct relationship between conventional labor force supply variables and short-run expenditure and financial portfolio variables.

Richard N. Rosett, in his study of working wives, found a positive relation between the level of debt and the labor force participation of wives. Since the data for his study were based on a cross section sample, a potential identification problem exists in that the wife's working may have reduced debt so that the extent of interdependence between *ex ante* debt and participation would be understated. Our model takes the more general view that there are a number of ways to move toward an overall portfolio equilibrium; one being increased debt, the other being increased earnings for the wife. The extent to which these two means of adjustment will be substituted for one another depends on the initial disequilibria.

Investigation of Proportional Adjustment

Very large disequilibria in the portfolio may present special adjustment problems to the consumer. This suggests that adjustment may not be proportional over all levels of disequilibria as is usually specified in stock adjustment models. To test for this possibility, higher order terms (quadratic and cubic) of the stock disequilibria were included in a reestimation of equation (2) in Table 3.

The introduction of both quadratic and cubic terms for *all* disequilibria does make some contribution to the formulation for all debtors ($F=8.2$, significant at the 1 percent level) but no single set of terms relating to a specific disequilibrium added significantly to the model. Using only quadratic terms for disequilibria measures also improved the results over the original formulation ($F=3.6$, significant at the 1 percent level). Again, no single term added significantly to the model. The largest single contribution ($F=1.9$) was made by the square of $[(P/Y)^{*}-(P/Y)_{-1}]$. The coefficients were 29.4 for $[(P/Y)^{*}-(P/Y)_{-1}]$ and .44 for the squared term. This suggests that a family making payments equal to 15 percent of disposable income and having an equilibrium payment ratio of 5 percent would reduce its debt by $249, less than double the reduction in debt it would make ($136) if its equilibrium payment level were 10 percent of income.[11] These results, though not particularly strong, are consistent with the hypothesis that large disequilibria are adjusted at a slower rate. For example, other studies have shown that payment difficulties are frequent among families with very high debt payment-income ratio.[12]

Overall, however, it appears that the assumption of proportionality is a reasonable approximation of the adjustment process in the sense that no other simplifying assumption appears to add much to the explanatory power of the model. Strict proportionality seems to be least appropriate for the disequilibria of relative payments but misspecification here has smaller consequences.

The Adjustment Process for Population Subgroups

Although a particular speed of adjustment may hold for the population as a whole, there are several characteristics of the adjustment process which suggest that the adjustment parameters may be very different for certain population subgroups. Three very important factors influencing the rate of adjustment are the overall liquidity of the portfolio, the scale of the portfolio, and its diversification. This suggests that the rate of adjustment coefficients may be variables across individuals rather than parameters of the system.

Arguments for economies of scale suggest that as the portfolio becomes larger, the average cost of adjustment for a given asset declines. But, as the *number* of different assets in the portfolio grows, the costs of coordinating adjustment rises. Thus it is possible that gains from scale economies for particular assets may be offset by the increased costs of coordinating a greater number of asset adjustments. Although we have no data relating the number of assets held to portfolio size, we do expect to observe slower adjustment toward equilibrium for a larger portfolio, given its liquidity.[13]

---

[11] 23.9(−10) and .44(100)

[12] A more complete discussion of the factors associated with repayment difficulties can be found in George Katona et al.

Given size, the more liquid the portfolio, the faster the consumer can move toward equilibrium since large deviations from desired levels in any asset can be met by adjustments in liquid assets. Moreover, if liquidity is thought of as a continuum from cash to assets like household durables, the more liquid the portfolio, the fewer discontinuities there are associated with the adjustment process. A formal demonstration of liquidity effects in the overall adjustment process would require a complete set of stock adjustment equations (see Brainard and Tobin).

A life cycle model of consumption suggests that there will be different points in the age and family composition of the household at which liquid assets are accumulated and at which depletions of liquid assets for the purchase of household durables will occur. Departing from the standard model (with its implied constant consumption path for an individual) and redefining income on a per capita basis, the expectation of increased family size (marriage, children) and a withdrawal of the wife from the labor force suggests that young spending units without children have an abnormally high income (relative to expected *per capita* levels). This life cycle pattern of asset adjustment is illustrated in Table 1. Young families with no children have large net holdings of liquid assets and lower equilibrium levels of household durables inventories than those families with young children present.

Because of these longer run considerations, differences in portfolio scale and liquidity are apparent across various subgroups of the population. To observe differences in the speed of adjustment among these groups, the sample was divided as follows:

1) family head under 35 years old with small children (age under 6)
2) family head under 35 years old, no small children
3) family head over 35 years old.

Following the format of Table 3 we have rerun our adjustment equation for these three subgroups and found considerable differences in adjustment speed. Results are shown in Table 4. An overall comparison of the estimates for those under 35 with small children and those over 35 suggests that the general model is valid across a wide variety of family (portfolio) structures. However, the ambiguity of the results for those under 35 with no children serves as a caveat against naive interpretations of such stock-adjustment models. This group exhibits very different adjustment coefficients. Given that our data are limited to annual observations, one possible interpretation is that for many of these families complete adjustment takes place *within* the time period of one year, and that such families move toward equilibrium very rapidly. For adjustment periods much longer than a single year (say for housing stocks), the annual model is valid because slower adjustment can be represented by a lower adjustment coefficient.[14] However, if periods of less than one year are sufficient for total adjustment, estimates from annual data would yield ambiguous results.

Even if adjustment is not completed within the sample period, the effect of

---

[13] It is assumed that the number of different assets held in the portfolio grows with portfolio size. One of the major "costs of adjustment" for a consumer is that of deciding whether a change in current income is temporary or is likely to persist since the two conditions imply different adjustments.

[14] Using aggregate housing data, Richard Muth estimated annual adjustment coefficients between .33 and .15 for housing stock adjustment which was lower than Chow's .55-.33 range for autos.

TABLE 4—STOCK ADJUSTMENT OF INSTALMENT CREDIT FOR DEBTORS BY AGE AND PRESENCE OF SMALL CHILDREN[a]

	(D^*-D_{-1})	(H^*-H_{-1})	(A^*-A_{-1})	$\left(\left(\frac{P}{Y}\right)^*-\left(\frac{P}{Y}\right)_{-1}\right)$	(W^*-W_{-1})	$\bar{R}^2$
Under 35 (all)	0.593 (.063)	.022 (.066)	.066 (.032)	14.73 (6.65)	−.041 (.041)	.298
Small Children	.713 (.084)	.058 (.087)	.055 (.044)	16.39 (8.99)	−.085 (.056)	.340
No Small Children	.392 (.091)	−.012 (.098)	.069 (.045)	16.79 (9.30)	.031 (.055)	.255
Over 35	.489 (.049)	.039 (.036)	.010 (.095)	14.76 (5.12)	−.043 (.022)	.281

[a] Standard errors in parentheses.

estimating adjustment parameters from data collected for time periods of a different length than the period of adjustment can bias estimates of the speed of adjustment (see Zellner (1968)). Applying his results to our findings suggests that estimates of the speed of adjustment are biased downward when actual adjustment takes place within the time period of observation. For example, suppose that the half year speed of adjustment coefficient were estimated to be b from panel data collected twice a year (instead of once a year as the data used in this study). Using *annual* data to estimate the rate of adjustment would yield b^*, related to b as follows: plim $b^*=b+b\ (1-b)$. This suggests that the estimate, b^*, from annual data consists of the half-year adjustment coefficient, b, plus b multiplied by the fraction of the disequilibrium remaining to be closed in the second half year. Thus, an adjustment coefficient estimated to be .39 (the direct adjustment parameter in the equation for families with no small children, Table 4) would imply a semiannual rate of .225, more than half the annual rate. Such bias, however, could not explain the dramatic differences in the coefficients across subgroups.

Disaggregation of the families under 35 into those with small children and those without small children has implications for other components of the adjustment process. Previous research has demonstrated that families with lower net asset (and overall wealth) positions face a consistently higher (marginal) cost of borrowing (see Stafford and Dunkelberg). As a result, these families should have a relatively higher equilibrium level of the wife's labor force participation (as well as other equilibrium portfolio differences) in attaining their portfolio balance. Our empirical results are consistent with this prediction and we do find that the point estimate of the cross-adjustment coefficient on the wife's labor supply (earnings) is largest for the subgroup with the greatest liquidity limitations—those under 35 with small children present.

Assuming that there is not a large share of either those under 35 with children and those over 35 who are adjusting *within* a year, then the results imply that (given liquidity) larger scale portfolios require a longer period for adjustment. Given a secularly growing consumer portfolio (see F. Thomas Juster), one would expect increasing demand for liquidity and/or a lengthening of the time period over which consumers make adjustments to erode any given disequilibrium.[15]

Such liquidity and scale differences may in part explain the differences between our findings and those of Watts and Tobin. They were working with data from a 1950 survey. The expansion of the consumer portfolio between 1950 and 1967 was roughly twofold. The value of household durables and automobiles rose from $21.9 billion in 1950 to $32.4 billion in 1962. The 1969 *Economic Report of the President* gives a figure for selected liquid assets held by the public of $271 billion in 1950 (in 1969 prices). Adjusting this liquid asset figure to 1962 prices and taking the ratio of liquid to real assets we have 250/285 = .87. Comparable data for 1962 yields a figure of .64 (367/577). Thus, in 1950, not only was the consumer portfolio much smaller but also much more liquid.

IV. Conclusions

This study has highlighted some of the aspects of consumer behavior with respect to management of the instalment debt in the portfolio. Although families tend to adjust their portfolios toward some desired level of instalment credit, there are wide variations about this level over time, a characteristic of behavior which can be revealed only through panel analysis.

Our findings of large year to year changes in credit obligations about some longer run level can be interpreted as evidence of potentially destabilizing effects of consumer credit. The implications for cyclical stability or instability of spending are not clear when the impacts of the variables included in the analysis are "added up" over individuals. In a time of rising aggregate demand, credit extensions will reduce the gap between desired and actual instalment credit and hence induce reductions or a slower expansion in credit during the subsequent period(s). However, employment potential for secondary earners is also rising and income is rising, diminishing the budgetary constraint of current obligations (the payment-income ratio). Income expectations (how large income will become and for how long) may be a prime determinant of debt incurrence, particularly for population subgroups that have a high rate of debt turnover. Expectations of short-run income variations were not explicitly included in this analysis.

Rising credit costs or limited "availability" (e.g., short maturities and higher payment to income ratios) will operate to dampen individual credit expansion. If debt and payment disequilibria operate in this fashion, there is an automatic tendency toward stability in the credit market. On the other hand, abrupt changes in statutory provisions concerning the length of maturities could affect payment to income ratios and result in dramatic shifts in credit outstanding and, by implication, in aggregate demand.

One question which we could not answer directly with the data in our study is the responsiveness of extensions to change in the borrowing rate. At a minimum, a higher rate for a given maturity and amount borrowed will result in a somewhat higher payment to income ratio. Insofar as we demonstrated an effect which would reduce subsequent extensions for families beyond some desired payment to income ratio, a higher credit cost would at least have some affect on account of such liquidity considerations in addition to its potential price effects.

Moreover, if consumers are in fact responsive to short-run changes in the cost of credit, a rise in credit cost (or limited "availability") could result in additional labor force participation of the wife. In a study of graduate student consumption (see Stafford), it was inferred that graduate students face high marginal borrowing costs and as a consequence, it was observed that graduate student wives work more than comparable wives whose husbands are not in graduate school. In this way students' wives served, in part, as a means for the student family to adjust to a stable flow of consumption and were an alternative to borrowing against future income.

---

[15] Although secular improvements in the banking and credit industry (increased velocity) would operate to reduce the time required for adjustment.

Apart from potential responses to rising credit costs, our study has shown that temporal variations in the use of instalment credit are also affected by disequilibria in other sectors of the household's finances. In particular, the wife's initial labor supply has a bearing on whether the family incurs new debt or encourages the wife to work more hours. These findings further imply that aggregate economic behavior may be better understood if work-leisure balance variables are integrated with the financial sector.

REFERENCES

G. S. Becker, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education,* New York 1964.

———, "A Theory of the Allocation of Time," *Econ. J.,* Sept. 1965, *75,* 493–517.

W. B. Bennett, "Consumption of Automobiles in the United States," *Amer. Econ. Rev.,* Sept. 1967, *57,* 841–49.

W. G. Bowen and T. A. Finegan, *The Economics of Labor Force Participation,* Princeton 1969.

W. C. Brainard and J. Tobin, "Pitfalls in Financial Model Building," *Amer. Econ. Rev. Proc.,* May 1968, *58,* 99–122.

G. C. Cain, "Unemployment and Labor Force Participation of Secondary Workers," *Ind. Labor Relat. Rev.,* Jan. 1967, *20,* 275–97.

G. C. Chow, "The Demand for Automobiles in the U.S.," in A. C. Harberger, ed., *The Demand for Durable Goods,* Chicago 1960.

M. David, *Family Composition and Consumption,* Amsterdam 1962.

J. Dusenberry, "The Portfolio Approach to the Demand for Money and Other Assets," *Rev. Econ. Statist.,* Feb. 1963, *45,* Supp., 9–24.

A. Enthoven, "Instalment Credit and Prosperity," *Amer. Econ. Rev.,* Dec. 1957, *47,* 913–29.

I. Fisher, *The Theory of Interest,* New York 1962.

M. Friedman, *A Theory of the Consumption Function,* Princeton 1957.

R. Holbrook and F. Stafford, "The Propensity to Consume Separate Types of Income: A Generalized Permanent Income Hypothesis," *Econometrica,* Jan. 1971, *38,* 1–21.

H. M. Hunter, "A Behavioral Model of the Long Run Growth of Aggregate Consumer Credit in the U.S.," *Rev. Econ. Statist.,* May 1966, *48,* 124–40.

J. Johnstone and R. Rivera, *Volunteers for Learning,* National Opinion Research Center, Chicago 1965.

F. T. Juster, *Household Capital Formation and Financing 1897–1962,* New York 1966.

G. Katona et al., *Survey of Consumer Finances, 1968,* Survey Research Center, Ann Arbor 1969.

J. B. Lansing, E. S. Maynes, and M. Kreinin, "Factors Associated with the Use of Consumer Credit," in *Consumer Instalment Credit,* Part II, Vol. I, Board of Governors of the Federal Reserve System, Washington 1957.

M. Miller and D. Orr, "A Model of the Demand for Money by Firms," *Quart. J. Econ.,* Aug. 1966, *80,* 413–36.

J. Mincer, "On-the-Job-Training: Costs, Returns and Some Implications," *J. Polit. Econ.,* Oct. 1962, *70,* 50–73.

———, "Labor Supply, Family Income, and Consumption," *Amer. Econ. Rev. Proc.,* May 1960, *50,* 574–83.

———, "Market Prices, Opportunity Costs and Income Effects," in C. F. Christ, ed., *Measurement in Economics,* Stanford 1963.

G. H. Moore and P. A. Klein, "The Quality of Consumer Instalment Credit," in Nat.

Bur. Econ. Res., *Stud. in Consumer Instalment Financing,* New York 1967.

E. Mueller and J. Lean, "The Savings Account as a Source for Financing Large Expenditures," *J. Finance,* Sept. 1967, *22,* 375–93.

R. Muth, "The Demand for Non-Farm Housing," in A. C. Harberger, ed., *The Demand for Durable Goods,* Chicago 1960.

J. Poapst and W. Waters, "Rates of Return on Consumer Durables," *J. Finance,* Dec. 1964, *19,* 673–77.

Y. B. Porath, "The Production of Human Capital and the Life Cycle of Earnings," *J. Polit. Econ.,* June 1967, *75,* 352–65.

R. N. Rosett, "Working Wives: An Econometric Study," in T. F. Dernberg et al., eds., *Studies in Household Economic Behavior,* New Haven 1958.

W. L. Smith, "Consumer Instalment Credit (A Review Article)," *Amer. Econ. Rev.,* Dec. 1957, *47,* 966–84.

F. P. Stafford, "The Income and Consumption of Graduate Students," unpublished doctoral dissertation, Univ. Chicago 1968.

——— **and W. Dunkelberg,** "The Cost of Financing Automobile Purchases," *Rev. Econ. Statist.,* Nov. 1969, *51,* 459–64.

D. B. Suits, "The Determinants of Consumer Expenditures: A Review of Present Knowledge," in D. B. Suits et al., eds., *Impacts of Monetary Policy, Commission on Money and Credit,* Englewood Cliffs 1963.

J. Tobin, "Consumer Debt and Spending: Some Evidence From a Survey," in *Consumer Installment Credit,* Part II, Vol. 1, Board of Governors of the Federal Reserve System, Washington 1957.

H. Watts and J. Tobin, "Consumer Expenditures and the Capital Account," in I. Friend and R. Jones, eds., *Proceedings of the Conference on Consumption and Savings,* Vol. II, Philadelphia 1960.

D. M. Wu, "Empirical Analysis of Household Durable Expenditures," unpublished doctoral dissertation, Univ. Wisconsin 1962.

A. Zellner, "Basic Elements of a General Model of Consumer Behavior," in I. Friend and R. Jones, eds., *Proceedings of the Conference on Consumption and Savings,* Vol. II, Philadelphia 1960.

———, "Note on the Effect of Temporal Aggregation on the Estimation of a Stock Adjustment Model," Univ. Chicago, unpublished manuscript 1968.

Council of Economic Advisors, *Economic Report of the President,* Washington Jan. 1969.

Estimation of Structure-Profit Relationships with Application to the Food Processing Sector

By BLAKE IMEL AND PETER HELMBERGER*

In putting the hypotheses of price theory to the test and in seeking to quantify relevant market structure-performance relationships, researchers have often employed multiple regression analysis with profit rate inserted as the dependent variable and with various measures of market structure serving as independent variables. While these studies represent varying levels of aggregation, the independent variables have generally been based on Census four-digit industries. In recent works, Marshall Hall and Leonard Weiss, David Kamerschen, and the Federal Trade Commission (*FTC*)[1] have sought to "explain" the variance among company profit rates. They use market-related variables that are the same for all firms in the industry (concentration for example) and firm-related variables which may take on unique values for and have unique impacts on individual companies. The size of the firm is an example of a firm-related variable. Another possible example, and one that may be impossible to measure, is the quality of a firm's management relative to that of its competitors.

A problem that has been observed in analysis of company profit rates, though it has appeared in more aggregative studies too, concerns heteroskedasticity. (Another problem which seems to have gone unnoticed is that of autocorrelation; but more on this later.) When classical least squares (*CLS*) is used to explain variation among profit rates of individual companies, the estimated residuals generally display heteroskedasticity. A few studies have attempted to deal with this problem through examination of residuals and through construction of *ad hoc* weighting schemes based on the sizes of companies.[2] Hall and Weiss, among others, have recognized that the problem might be explicable largely in terms of diversification.

In what follows, an econometric model is developed which suggests that in analysis of company profit data, generalized least squares (*GLS*) is the appropriate estimation procedure when there is interest in using statistical tests. By implication, the *t*-ratios for estimates based on

---

* Agricultural economist, Economic Research Service, U.S. Department of Agriculture; professor of agricultural economics, University of Wisconsin, Madison, respectively. We are indebted to Michael Behr who contributed importantly to many phases of the project, particularly in planning and implementing the field survey and in locating secondary data. We wish to thank Rueben Buse, Arthur Goldberger, A.C. Johnson, Jr., and especially Willard Mueller for their helpful comments. The research was supported by the College of Agricultural and Life Sciences, University of Wisconsin, the Giannini Foundation, University of California, Berkeley, and the Economic Research Service, USDA, and is reported in detail in a Ph.D. dissertation by Imel. The authors take responsibility for any errors.

[1] The *FTC* study is similar to our own with the following major exceptions: First, though both studies focus on about the same sample of companies, the *FTC* uses data for the period 1949–53; our data are for 1959–67. Second, while some variables are common to both researches, others are only similar. Third, and as explained below, our estimation procedures differ markedly from those used by the *FTC*.

[2] See Hall and Weiss and the *FTC*. As noted, heteroskedasticity may also be a problem in more aggregative studies (see William Comanor and Thomas Wilson). The theoretical model presented below can be extended to cover the aggregative studies.

classical least squares lack meaning. Problems associated with heteroskedasticity and autocorrelation are anticipated through a priori specification with diversification playing a key role in the formulation.

The results from an empirical application of the model, given in the latter part of this paper, should be of interest in their own right. The application involves estimation of structure-profit relationships using data for 99 food and tobacco manufacturing companies and their markets. The sensitivity of regression results to different statistical specifications and different industry definitions is analyzed. We conclude that concentration and product differentiation are important variables in explaining variation among profit rates of sample companies. Estimated coefficients for plant scale barriers to entry, absolute firm size, market growth and diversification, had the "wrong" signs and/or very low *t*-ratios. Moreover, *CLS* estimation led to inflated *t*-ratios. Statistical results improved by the usual standards as industry definitions became less gross and approached theoretic industries more closely.

I. The Econometric Model

It is instructive to begin by considering a world which is much simpler than the one we know in that all firms are completely specialized, single-product firms. Suppose further that a researcher selects a random sample of firms which, fortuitously, just happens to contain no more than one firm in any one industry. With the resulting survey data, *CLS* regression might be employed to estimate the parameters of the following equation:

$$R_{ij} = a + bM_j + cE_{ij} + u_{ij} \tag{1}$$

where

R_{ij} = the ratio of economic profit to sales (or some measure of capital) of the ith firm operating in the jth industry,

M_j = a market-related structural variable, such as a concentration ratio, which is common to all firms in the jth industry,

E_{ij} = a firm related variable, such as quality of management (or the degree of product differentiation) relative to the industry average for the ith firm in the jth industry,

u_{ij} = an error term resulting from the omission of both firm and market explanatory variables.[3]

If $E(u_{ij}) = 0$; $E(u_{ij}u_{kh}) = \sigma_u^2$ when both $i = k$ and $j = h$, and 0 otherwise; and if the error term is uncorrelated with the included independent variables, the estimates of the parameters a, b, and c will have certain desirable properties. Moreover, statistical tests of significance are available if the error term, u_{ij}, is normally distributed. All of these assumptions concerning the error component appear reasonable under these controlled conditions. However, there are potential difficulties when the sample contains more than one firm from any one industry and where some firms are diversified. Each of these complications is taken up in turn.

It is, of course, quite possible that a random sample of specialized firms could contain a number of firms from any one industry. To see the resulting implications for the properties of the error term, it is instructive to split that term, u_{ij}, into two components:

$$u_{ij} = e_{ij} + m_j \tag{2}$$

where e_{ij} is that portion of the total error term attributable to omission of firm-related variables and m_j equals the error due to the omission of market-related variables. In light of the available data and the complexity of the real world, it appears almost certain that many independent variables—some market-related and some firm-related—will usually be omitted in empirical work.

---

[3] An alternative rationalization of the errors in the equation asserts an element of inherent indeterminancy in economic relationships dependent upon human behavior. A distinction between the two rationalizations is not always practical (see J. Johnston, pp. 5–6), and we adopt that given in the text above.

The decomposition of the error term into two components is consistent with the previous formulation if the error components have the following sufficient properties:

$$\begin{aligned} E(e_{ij}) &= E(m_j) = 0 \\ E(e_{ij}e_{kh}) &= 0 \quad \text{where} \quad \text{either } i \neq k \\ &\qquad\qquad\qquad\qquad \text{or } \ j \neq h \\ E(e_{ij}m_h) &= 0 \quad \text{where} \quad j \neq h \\ E(m_j m_h) &= 0 \quad \text{where} \quad j \neq h \end{aligned}$$

Furthermore, by definition, omission of market-related variables leads to an m_j which is identical for all firms in an industry so that $\mathrm{Cov}(e_{ij}, m_j)=0$. Appealing to the central limits theorem, it is assumed that both m_j and e_{ij} are normally distributed with a finite variance.

With these specifications in mind, let us return to equation (1) where j ranges over $1 \ldots m$ markets and i may range over $1 \ldots n$ firms although it is to be understood that the number of firms in each industry may vary. The variance-covariance matrix of u_{ij}, hereafter referred to as the *omega* matrix, will be examined in more detail. Diagonal terms are given by:

$$(3) \qquad \begin{aligned} \mathrm{Var}(u_{ij}) &= \mathrm{Var}(e_{ij}) + \mathrm{Var}(m_j) \\ &\quad + 2\,\mathrm{Cov}(e_{ij}, m_j) \\ &= \sigma_e^2 + \sigma_m^2 + 0 \end{aligned}$$

It can further be shown, using the assumptions laid out above, that:

$$(4) \quad \mathrm{Cov}(u_{ij}, u_{kh}) = \begin{cases} \sigma_m^2, & \text{if } j = h \\ 0, & \text{if } j \neq h \end{cases}$$

If two firms operate in the same market, they have the same m_j submerged in their error terms; the expected cross-product of the two m_j's is σ_m^2. The expectation of the remainder of the cross-products, including the cross-products of the m_j's for two firms not in the same market, is assumed to be zero. The emergence of a positive off-diagonal element (wherever two firms are in the same market) in the *omega* matrix means that *CLS* is no longer appropriate.

We turn next to a somewhat more complex world in which at least some of the firms may be diversified. It is assumed that a diversified firm is nothing more or less than an aggregate of single-product firms (each of which will be referred to as a division) so that all our assumptions about the relationships of the various components of the error term still apply. Obviously, this is but an approximation of the real world and we shall have more to say about a more realistic formulation at a later point.

Consider, then, the profit function of the gth and kth firms given below where P equals the level of profit and S the level of sales:

$$(5) \quad P_{gj} = (a + bM_j + cE_{gj} + e_{gj} + m_j)S_{gj}$$

$$(6) \quad P_{kh} = (a + bM_h + cE_{kh} + e_{kh} + m_h)S_{kh}$$

The ith diversified firm, on the other hand, which consists of these two "firms" or divisions would have the following profit rate function:

$$(7) \qquad \begin{aligned} R_{id} = a &+ b(D_{ij}M_j + D_{ih}M_h) \\ &+ c(D_{ij}E_{gj} + D_{ih}E_{kh}) \\ &+ D_{ij}(e_{gj} + m_j) + D_{ih}(e_{kh} + m_h) \end{aligned}$$

where

R_{id} = the ratio of total profit, P_{id}, to total sales, S_{id}, for the ith diversified firm,

D_{ij} = the share of the diversified firm's total sales accounted for by its sales in the jth market,

D_{ih} = the share of the diversified firm's total sales accounted for by its sales in the hth market.

Equation (7) shows the necessity of using weighted independent variables in regression analysis of diversified firms; a suggestion which is scarcely new. Note further, that profit was expressed as a percentage of sales. Had profit been expressed relative to assets (equity), the appropriate weights would have been changed correspondingly to the percentages of assets (equity) accounted for by the various divisions. As a practical matter, however, it would appear very difficult to allocate assets and impossible to allocate net worth among the various divisions of a diversified firm.

The error term from equation (7) is rather complex and merits further attention. Let f_{id} equal the total error term for the ith diversified firm. It can be shown on the basis of previous assumptions that:

$$E(f_{id}) = 0 \tag{8}$$

$$\mathrm{Var}(f_{id}) = (D_{ij}^2 + D_{ih}^2)\sigma_u^2 \tag{9}$$

These results can be generalized for any diversified firm. For the kth firm, say, that operates in all m markets:

$$\mathrm{Var}(f_{kd}) = \sum_{j=1}^{m} D_{kj}^2\sigma_u^2 \tag{10}$$

Turning to covariance terms, if any two firms, k and i, whether diversified or not, have no market in common, the covariance of their error terms is zero. If the two firms, one of which is specialized, have one market in common, say market j, then:

$$\mathrm{Cov}(u_{kj}, f_{id}) = D_{ij}\sigma_m^2 \tag{11}$$

If the two firms are each engaged in all m markets:

$$\mathrm{Cov}(f_{kd}, f_{id}) = \sum_{j=1}^{m} D_{kj}D_{ij}\sigma_m^2 \tag{12}$$

In general, the covariance term for any two firms, i and j, that have at least one market in common is given by $\alpha_{ij}\sigma_m^2$ where the coefficient α_{ij} equals the sum of the cross-products of the D_{ij}'s from the firms' common markets. Note that α_{ij} also has a maximum value of one which occurs when both firms are specialized in the same market. It is less than one otherwise and, of course, may vary among various pairs of firms, depending upon their patterns of diversification.

We are now in a position to sum up. Realizing that some firms are specialized and others diversified to varying degrees and that a sample of firms might include a varying number of firms in various industries, the composition of the *omega* matrix will depend on which firms are included in the sample. If, for any sample, we factor σ_u^2 out of the *omega* matrix, the diagonal elements will consist of Herfindahl measures of product diversification, i.e., the relative shares of the company's total sales accounted for by various markets in which it operates are simply squared and summed. If the firm is completely specialized, the appropriate diagonal element will be one. If the firm is diversified, the element is less than one. Any off-diagonal element for two firms that share no market will be zero. Any off-diagonal element for two firms that share at least one market, the ith and jth firms, say, will be given by:

$$\alpha_{ij}K \tag{13}$$

where $K = \sigma_m^2/\sigma_u^2$ and is hereafter referred to as the *omega* ratio.

One may correctly accept *CLS* as an appropriate means of estimating the parameters of equation (1) in the hypothetical case of the single-product enterprise with a fortuitous sample of firms, one from each industry. If some sample firms are diversified and/or the sample contains more than one firm from any one industry, then it appears that *GLS* is the appropriate estimation procedure. In this latter case, the *CLS* estimates of the regression coefficients are unbiased though they are not minimum variance (best) and the estimated standard errors of the coefficients are meaningless (see Arthur Goldberger, pp. 238–39).

Unfortunately, one component of the variance-covariance matrix—the *omega* ratio—is not known a priori and appears to be beyond estimation. While this precludes straightforward application of the proposed model, all is not lost. If the breakdown of sales for each sample company according to the markets in which it operates were available, it would be possible to calculate the diagonal elements and the α_{ij}'s and use sensitivity analysis to investigate the consequences of setting the *omega* ratio at different levels. This is the plan for the empirical work reported on below. More specifically, we collected the necessary data in a field survey and computed all elements of the *omega* matrix except the *omega* ratio, K. That ratio was set at three alternative levels within the acceptable range of zero to one. The *omega* matrix is singular if the ratio equals 1.0 and we chose the values 0.0, 0.5, and 0.8.

The conception of the diversified firm as a mere aggregate of single-product firms is not very convincing in some respects. The main problem seems to be that one might not expect the e_{ij}'s associated with the various divisions of a diversified firm to be completely independent if some of the omitted firm variables pertain to overall operations of the firm. An example of such an omitted variable might be the quality of top-level management.

This problem can be treated by decomposing the error term for the diversified firm into three components, each attributable to the omission of a group of variables related to the firm's profitability. One group of variables relate to the nature of markets in which the firm operates, another to the divisions within the firm such as the quality of divisional management, and a third to the factors affecting overall firm operations. The variance-covariance matrix derived for this alternative model is similar to the one outlined above. The off-diagonal elements are, in fact, the same. The diagonal elements are altered by the addition of a constant term which, of course, decreases the degree of heteroskedasticity. Rank is preserved perfectly. The number of unknown parameters is increased, however, and for this reason the model is not considered further here.

II. Structure-Profit Relationships in the Food Processing Sector

Implementation of the estimation procedure suggested in the above formulation calls for data that are unavailable from secondary sources. In time, especially with the current pressure for corporate reporting on a divisional basis, this situation might change. In the interim, the usefulness of the above econometric model and the prospects for meaningful replication (mainly in other sectors) of our research results will hinge on the ability of researchers to collect a considerable amount of data which firms will be unwilling to make available except on a confidential basis. In this regard, our research on structure-profit relationships in the food sector may be viewed as a test case. We conclude that it may be quite possible for researchers to supplement secondary sources with the necessary primary data.

A field survey of companies in the food sector, defined here to include tobacco, was conducted in the following manner: from *Fortune Magazine's* list of the 1000 largest *U.S.* manufacturing corporations in 1965, all those firms in the food sector were identified. Some companies not primarily engaged in the food sector were excluded because the collection of information on their nonfood operations was deemed to be too costly. Other companies refused to cooperate in the survey. Sufficient information was collected to allow the inclusion of 99 companies in the analysis. Primary data were obtained through personal interviews, phone calls, and mailed questionnaires; accuracy was checked wherever possible with data from published sources. More details on the nature of the data used in the analysis will be given in the subsequent discussion of the variables. Clearly, the sample of firms analyzed was not randomly selected. There was a definite bias toward large firms and in favor of industries where the average size of firm is large. The combined total of 1963 food sales of sample companies was nearly $30 billion, a figure which represents over 40 percent of the sales of the food sector for that year.

Because little work has previously gone into the framing of industry definitions, apart from culling and minor alteration of Census groupings, a sensitivity analysis was made on the basis of three alternative sets of industry definitions within the food sector. One set is simply the often used Census four-digit industries, except that cane and beet sugar have been combined into one industry and likewise for the three four-digit oil crushing groups. A second set, and the one receiving the most attention in subsequent analysis, represents a substantial effort to define industries within the food sector which meet as closely as possible economic criteria. Information collected in the survey and from knowledgeable academic personnel and various secondary sources was useful in this effort. No particular level of Census classification was adhered to as four-digit, five-digit, and seven-digit groups were merged, split, etc. Census data (see U.S. Department of Commerce; U.S. Senate Subcommittee on Antitrust and Monopoly) were of considerable value in the computation of various statistics for the new theoretic groups, however, and they are, therefore, referred to as the reconstructed census industries. Subjective judgments were rendered at nearly every stage of the exercise, though it should be emphasized that one cannot avoid making such judgments, implicitly at least, by simply using Census industries.[4] The third set of industries consists merely of four-digit and five-digit Census groups that were chosen to correspond as closely as possible to reconstructed Census industries.

There are many differences, of course, between the three sets of industry definitions; two merit some mention. First, all Census concentration indices are calculated on the basis of national markets. The appropriateness of this procedure may vary somewhat among industries, but it is a flagrant error in the case of soft-baked goods and fresh dairy products. After some deliberation and consideration of various data sources (Alden Manchester, pp. 48–51; National Commission on Food Marketing, pp. 52–53), it was decided to use states to approximate markets for soft-baked goods and federal milk marketing order areas as milk markets in the reconstructed census definitions.[5] Second, as is well known, the Census four-digit groups tend to follow similarities in process, often ignoring considerations of buyer substitution. In reconstructing the Census industries, we have tried to give more weight to substitutability in the eyes of buyers. The reconstructed Census and Census four-five-digit industries are listed in Tables 1 and 2 of the Appendix.

---

[4] For a discussion and an application of economic criteria for industry delineation, see Carl Kaysen and D. Turner, pp. 295–324.

[5] A similar adaptation of regional markets was used by the *FTC* study, p. 9.

1. The Variables

There is some controversy over which profit rate is the most appropriate for use as a dependent variable in analyses of this sort; good arguments can be adduced in support of any one of several of them (see Kamerschen; Joe Bain, pp. 386–96; George Stigler, pp. 51, 123–25).

In addition, the various profit rates tend to be correlated. As noted above, an endearing aspect of profit over sales is its internal consistency with allocating the activities of the diversified firms among markets on the basis of sales. For these reasons we will simply summarize our main statistical results using the rate of excess profit on sales. Specifically, the dependent variable is

R_i = total profits, after tax and after a deduction for implicit returns to equity capital (0.05 times year-end equity), divided by total sales, ith company, for the period 1959–67.[6]

Averages over time are employed to avoid strictly short-run phenomena. Accounting profits have been corrected for the omission of implicit rates of return on equity; shortcomings due to price level changes, differing time patterns of asset acquisition, arbitrary methods of computing depreciation, and other anomalies arising out of the curious ways of accountants are merely recognized. For the diversified firm, each of the following independent variables except company size and diversification is weighted using division sales as weights. Sales from foreign operations are ignored in computing weighted variables in that structures of foreign and domestic markets are assumed to be the same. (No attempt was made to collect data on the structures of foreign markets.) A further complication concerns intermediate goods of vertically integrated firms that could have been, but were not, sold in vertically related markets. Survey information on the "sales value" of these intermediate products allowed treating them as we have treated sales from the divisions of any other diversified firm. Finally, some companies are diversified because, in part, they sell in geographically separated markets. This was true of companies producing dairy and/or soft-baked goods. The dairy and soft-baked sales of any company were allocated among the various regional markets on the basis of employment.[7] The assumption here was that the ratio of "dairy employment" in a local market to total dairy employment of the company equaled the ratio of dairy sales in the local market to total company dairy sales. The same assumption was made in the case of soft-baked goods.

The major independent variables included in our analysis are listed below and some comments regarding theoretical justification are given thereafter:

M_{1j} = sales of product j by the four largest sellers of product j divided by total market sales of product j, 1963.[8]

---

[6] All financial data were taken from *Moody's Industrial Manual*. For some companies financial data were not available for the full nine-year period. In no case, however, does the weighted average include fewer than four years. In all cases the firm-related independent variables are based on the years for which financial data are available.

[7] Company sales by geographic region were not collected in the field survey but we were fortunate to secure a secondary source which listed location, four-digit census classification, and number of production employees for each plant for firms selling in these geographic markets. These data were purchased by the USDA from McGraw-Hill and were made available to us for this study.

[8] Concentration ratios for the Census four- and Census four-five-digit industries were taken from the census sources. The concentration ratios for the reconstructed census industries were compiled from information on market shares of sample firms and their competitors and from other sources such as the *FTC* reports, trade publications, corporate reports, knowledgeable industry and academic personnel, and the like. Thirty of the fifty-four reconstructed industries are the same as or may be identified closely with Census four- and five-digit

E_{1i} = estimated total advertising expenditures by the ith company in traceable mass media (all TV, radio, magazine, newspaper, and billboards) divided by the total ith company sales for the nine-year period, 1959–67.[9]

E_{2i} = estimated total research and development (R and D) expenditures by the ith company divided by total company sales, ith company, for the nine-year period, 1959–67.[10]

M_{2j} = estimated sales of the minimum efficient size plant divided by total industry sales, jth industry, 1963.[11]

E_{3i} = sales of product j by the ith company divided by the sales of product j by the four largest sellers of product j, 1963.

E_{4i} = average of year-end total assets, in ten-thousands, ith firm, 1959–67.

E'_{4i} = reciprocal of the log of average year-end total assets, in thousands, ith firm, 1959–67.

M_{3j} = the compound growth rate for market j, interpolated from total growth index for market j, 1954–63.

E_{5i} = the percentage of sales by firm i which is not classified in its primary reconstructed census industry.

E'_{5i} = the percentage of sales by firm i which is not classified in its primary four-digit census industry.

E''_{5i} = the percentage of sales by firm i which is not classified in its primary three-digit census group.

The ability to see the rewards associated with collusion and the willingness to participate in and enforce collusive agreements would appear to increase as concentration increases. On this well-worn argument, we include M_{1j} as appropriately computed for each of the three industry definitions. The expected sign of the parameter is positive.

A firm with a highly differentiated product might be insulated from competition on the part of both established and potential sellers. We include variables E_{1i} and E_{2i} to measure the magnitudes of two potentially major sources of product differentiation. Ideally, we would like to include both industry averages and division (or firm) deviations from these averages but the necessary data were not available.

---

groupings. In these cases our concentration ratios never vary by more than five percentage points above or below the corresponding census ratios. For more detail on all of the variables see Imel.

[9] Estimates of expenditures on traceable mass media—network and spot television, newspapers, magazines, newspaper supplements, radio, and outdoor advertising—were compiled from several secondary sources (see Bureau of Advertising of American Newspaper Publishers Association; Institute of Outdoor Advertising, Inc.; Leading National Advertisers, Inc.; Radio Expenditure Reports; and Television Bureau of Advertising, Inc.) for the years 1956, 1960, 1967. Outdoor advertising was not included for 1956 and 1960, nor was radio for 1956 (the relative magnitudes of these latter expenditures are very small). The data are subject to several shortcomings, the most objectionable being the understatement of expenditures of small regional firms. In the field survey an attempt was made to correct for this and other deficiencies as well as to ascertain the validity of linear extrapolation over the periods for which the secondary data had not been secured. The estimates were adjusted with aid from respondents. No attempts were made to collect data in nonmedia promotional activities.

[10] Estimates of R and D expenditures were collected in the field survey. In interviews and in correspondence, R and D expenditures were defined as those "incurred to create new products and processes or improve old ones." An effort was made to use terminology close to that used by the National Science Foundation (p. 106). As a check on accuracy, the coefficient of correlation between R and D and number of patents secured (1959–67), both expressed as a percentage of sales, was .94. (Annual data on patents secured by company are published by the U. S. Patent Office.)

[11] Estimates of minimum efficient size plants are taken from the National Commission on Food Marketing with the following exceptions: cigars, cigarettes, distilled liquors, and malt liquors. Estimates for these latter industries are based on survey data while the Commission's estimates are based on analyzing census data using the survivor technique. Estimates are available for Census four-digit industries only.

Advertising expenditures have been emphasized by Bain, p. 227, and others as a major source of product differentiation.[12] The arguments would appear to be especially true in the food sector where other determinants of product differentiation would appear to be generally unimportant, e.g., product durability and complexity and integration of dealer-retail organizations.

Research and development expenditures may give rise to product differentiation based on real product differences in terms of either hard-to-imitate product superiority or a continuing stream of new modifications and improvements.[13] Research and development programs may also foster process advantages which allow the firm to produce at a relatively low cost. Barriers or advantages stemming from R and D or advertising may become tied in with considerations of scale and capital requirements.

If optimum firm size is large relative to the industry and diseconomies of a less than optimum size operation are large, the potential entrant must either depress the market price and/or incur relatively higher costs. The variable M_{2j} is included on the rationale that plant size alone may, to some extent, be important as a barrier. It is fully realized that economies (diseconomies) associated with firm size and differences in price elasticities of demand (see Franco Modigliani) might render the estimates meaningless.

The *FTC*, p. 10, has devised a measure of relative market share, which they say signals the existence of product differentiation and possible economies of scale. This variable, E_{3i}, more appropriately called relative firm size, is included in our analysis because of its statistical significance in the *FTC* study rather than on strong theoretical grounds. The variable may be a proxy for progressiveness and good management and a number of other factors in addition to product differentiation.

William Baumol, p. 33, argues that size of firm alone may contribute to a higher rate of return because increased money capital may put firms on a higher echelon of imperfectly competing capital groups. To test this hypothesis E_{4i} and E'_{4i} are used. The latter, used by Hall and Weiss and the *FTC*, allows for nonlinearities in the functional form.

There are two theories concerning the influence of growth of market demand on profit rates. Some theorists (see Bain, p. 301; Richard Caves, pp. 30–31; and Carl Kaysen and D. Turner, p. 105) have hypothesized a negative relationship on the argument that collusive agreements are harder to maintain when demand is increasing rapidly. Others (see Hall and Weiss, Kamerschen) have hypothesized a positive relationship presumably because of errors in demand expectations and a lag in supply response. Both theories might be operative making it impossible to predict a net effect. Growth in market sales is used here to measure growth in demand. The fact that similar measures have been employed elsewhere does not diminish its abject crudeness.

The hypothesis that diversification increases a firm's market power and, therefore, its profitability is tested using E_{5i} and two variants thereof (see Corwin

---

[12] Comanor and Wilson, p. 234, argue further that to some extent, advertising expenditures may also be taken as a symptom of differentiation.

[13] Our treatment of R and D activities as a source of product differentiation is not new. Oliver Williamson, for one, has noted, p. 28, that in the long run, research and development expenditures as well as selling costs are relevant in the determination of product differentiation barriers. It should also be noted that as advertising and R and D expenditures are firm-related variables they should, theoretically, be collected on a divisional basis (for the diversified firm) and then aggregated by weighting. A moments reflection will show, however, that this recommended procedure results in the same ratios as described above, which are then implicitly "weighted divisional variables."

Edwards). Our test, of course, focuses only on static relationships which may not be important in this case. Even at that, the variable may be a bit naive in view of the complexities of conglomerate theory.

2. *Empirical Results*

The *omega* ratio was set at the levels of 0.0, 0.5, and 0.8 in order to allow *GLS* estimation. Estimates obtained from *CLS* are also shown for purposes of comparison. To conserve space, major results are given in tables with little additional comment.

The estimated equations shown in Table 1 incorporate the independent variables used to measure concentration and product differentiation. Turning to the reconstructed census definitions, the coefficients are all positive and in agreement with a priori expectations. The estimates are all unbiased, in the absence of other errors previously noted, but vary markedly from one formulation to the next. We submit that the researcher using *CLS* estimation, unaware of the limitations noted above, would be immensely gratified by the *t*-ratios. Yet these ratios lack meaning and the *t*-ratios in the *GLS* formulations tend to fall as the value of the *omega* ratio is increased. Importantly, however, the *t*-ratios are generally robust in all formulations and, we feel, offer good evidence that the relevant population parameters are, in fact, positive. The evidence is especially compelling for concentration.[14]

Regarding sensitivity of results to industry definition, the estimated coefficient and *t*-ratio for the one market-related variable—concentration—tend generally to be lower the greater are the departures of industry definitions from those judged to fit most closely economic criteria, viz., the reconstructed census definitions.[15] The pattern is not unexpected in that market variables could contain larger errors the less "theoretically accurate" the industry definition. The estimated coefficients and *t*-ratios for the firm variables, advertising and *R* and *D*, are generally greater the further are the departures of the industry definitions from the reconstructed census definitions. This pattern is not unexpected. The values of these two variables are not affected by the change in industry definition but their correlation with concentration is less the poorer the industry definition. Finally, the patterns of change in the estimates and *t*-ratios as the *omega* ratio is scaled upward seem to be about the same for all three sets of industry definitions.[16]

---

[14] Under the assumptions set forth in the statistical model, the variance of the error term associated with diversified firms is directly proportional to the index of diversification, $\sum D_{ij}^2$. We would expect, therefore, that plots of the *CLS* calculated residuals would exhibit a wider scatter as this index of diversification rises. This is, of course, a case of heteroskedasticity. The expected pattern was in fact observed, especially when comparing the two extremities of the range. Equivalent plots of the calculated residuals from the *GLS* models do not display heteroskedasticity. The sample was also divided into quartiles by the values of the firms' $\sum D_{ij}^2$'s and the average squared residual computed for each quartile. For *CLS*, the relative values of the group averages were about what would be predicted by the relative values of the group average $\sum D_{ij}^2$, the theoretical index. Again in *GLS* no pattern was discernible.

[15] The variables for the reconstructed census regional markets—soft-baked and fresh dairy products—were recalculated on a national basis and the equations of Table 1 reestimated. While the estimated coefficient and *t*-ratio of the market-related variable fell as one would expect, they were still quite substantial and in excess of the corresponding estimates from Census definitions. It appears that our conclusions are not particularly sensitive to the use of these particular regional market definitions.

[16] When total returns (profit plus interest, see Stigler, pp. 51, 123–25) to total assets and profit over net worth are used as alternative dependent variables, the signs in the above equations are the same and the same general patterns for *t*-ratios and coefficients occur with respect to changes both in the market definitions and in the *omega* ratio. Though the *t*-ratios are somewhat lower, conclusions reached above are not substantially altered. Concerning the *t*-ratios, we have noted that the sales weighting scheme adopted here is inappropriate for these two measures of profits, and explicitly introduces errors in the variables.

TABLE 1—NINETY-NINE FOOD SECTOR COMPANIES: ESTIMATED REGRESSION EQUATIONS RELATING PROFIT OVER SALES (R_i) TO THREE EXPLANATORY VARIABLES USING CLASSICAL LEAST SQUARES (*CLS*) AND USING GENERALIZED LEAST SQUARES (*GLS*) WITH THREE ALTERNATIVE SPECIFICATIONS OF THE VARIANCE-COVARIANCE MATRIX OF THE ERROR TERM

		Regression Coefficients With *t*-ratios in Italics			
Estimation Procedure	Intercept	Four-Firm Concentration Ratio (M_{1j})	Advertising Over Sales (E_{1i})	Research and Development Expenditures Over Sales (E_{2i})	Coefficient of Multiple Determination (R^2)
		Reconstructed Census Industries			
CLS	−.0084	.0376	.1977	.6706	.459
		3.44	*3.34*	*3.72*	
GLS: 0.0[i]	−.0111	.0393	.2000	.6751	.695
		3.48	*3.31*	*3.30*	
GLS: 0.5[i]	−.0164	.0533	.1433	.5511	.428
		3.01	*1.86*	*2.14*	
GLS: 0.8[i]	−.0280	.0747	.1119	.3761	.250
		3.10	*1.29*	*1.20*	
		Four-Five-Digit Industries			
CLS	−.0019	.0256	.2541	.8323	.429
		2.49	*4.59*	*4.81*	
GLS: 0.0[i]	−.0036	.0222	.2488	.7712	.674
		2.34	*4.17*	*3.71*	
GLS: 0.5[i]	−.0030	.0311	.1830	.7191	.400
		2.07	*2.38*	*2.85*	
GLS: 0.8[i]	−.0064	.0394	.1407	.6117	.202
		1.83	*1.67*	*1.99*	
		Four-Digit Industries			
CLS	−.0008	.0189	.2760	.9000	.411
		1.77	*5.02*	*5.22*	
GLS: 0.0[i]	−.0015	.0218	.2643	.8407	.670
		2.04	*4.53*	*4.13*	
GLS: 0.5[i]	−.0026	.0327	.1948	.7643	.399
		1.96	*2.58*	*3.05*	
GLS: 0.8[i]	−.0062	.0420	.1528	.6573	.199
		1.73	*1.73*	*2.15*	

[i] The *omega* ratio equals the value indicated. See text for interpretation.

Numerous equations, incorporating the measures of relative and/or absolute firm size variables were estimated using both *CLS* and *GLS*. A cross section of representative results is given in Table 2. While the coefficient of the relative size variable is rather small, it has the expected sign and the *t*-ratio is quite substantial. The result is typical for all *CLS* and *GLS* formulations and for all industry definitions. Thus, the *FTC* hypothesis is given strong support. We continue to have some reservations, however, about the original economic rationale for including this variable in the first place.

Of the two alternative measures of absolute size, E_{4i}, the linear form, consistently provided the stronger estimates. When the only other independent variable is the four-firm concentration ratio, the coefficients of both size variables have the correct signs—assets with a fairly high *t*-

TABLE 2—NINETY-NINE FOOD SECTOR COMPANIES: ESTIMATED REGRESSION EQUATIONS RELATING PROFIT OVER SALES (R_i) TO VARIOUS EXPLANATORY VARIABLES USING GENERALIZED LEAST SQUARES (*GLS*) WITH *Omega* RATIO EQUAL TO 0.5, RECONSTRUCTED CENSUS INDUSTRIES

	Regression Coefficients With *t*-ratios in Italics						
Intercept	Four-Firm Concentration Ratio (M_{1j})	Advertising Over Sales (E_{1i})	Research and Development Expenditures (E_{2i})	Relative Firm Size (E_{3i})	Assets (E_{4i})	Inverse of Log Assets (E'_{4i})	Coefficient of Multiple Determination (R^2)
−.0231	.0514	.1092	.4865	.0269			.469
	3.00	*1.44*	*1.94*	*2.71*			
.0027	.0706				.0018		.405
	4.45				*1.85*		
−.0171	.0512	.1106	.5416		.0014		.440
	2.89	*1.38*	*2.11*		*1.41*		
−.0229	.0508	.0995	.4876	.0250	.0005		.471
	2.94	*1.27*	*1.94*	*2.33*	*0.50*		
−.0226	.0727					−.2696	.394
	4.54					*1.27*	
.0012	.0516	.1265	.5472			−.1928	.433
	2.89	*1.59*	*2.12*			*0.90*	
−.0279	.0518	.1122	.4851	.0279		.0492	.470
	2.99	*1.45*	*1.93*	*2.54*		*0.21*	

ratio. Though this tends to support the Baumol hypothesis, the significance of the coefficients is substantially weaker with the addition of other independent variables. In fact, E'_{4i} has the wrong sign when relative firm size is included in the equation. The additional variables are positively correlated, to varying degrees, with either measure of size. In our judgment, the overall results for our sample lend support to the Baumol hypothesis in a limited context only. That is to say, when product differentiation and other factors which may be associated with firm size are taken into account, absolute firm size offers little additional explanatory power.

Plant Scale Barriers, Growth, and Diversification

The formulations reported in Table 1 were modified by adding, one at a time, each of the following three independent variables: plant scale barrier to entry, M_{2j}; market growth, M_{3j}; and diversification, with three alternative measures employed. Regarding M_{2j}, the estimated coefficient was always positive, as expected, but the *t*-ratios were not very impressive using *CLS* estimation and were even lower in the various *GLS* formulations. Using reconstructed census industry definitions, for example, and setting the *omega* ratio equal to 0.5, the estimated coefficient equaled .3532 with a *t*-ratio equal to 1.07. Interestingly, though we don't know quite what to make of it, the inclusion of M_{2j} plays havoc with the *t*-ratios for concentration when Census four-digit industries are used. Similar results have been reported by Comanor and Wilson who suggest that multicolinearity is partly to blame.

The estimated coefficients for M_{3j} were positive with small *t*-ratios using *CLS* and *GLS* with zero off-diagonal elements; the estimates for the initial three independent variables were not greatly affected. The estimated coefficients were negative with *t*-ratios in excess of 3.0, however, using *GLS* with non-zero off-diagonal elements. The results were obviously highly sensitive to the statistical procedure used and, on balance, offer little support for the hypothesis that market growth has a systematic effect on company profit rates.

Estimates did not support the hypothesis that the extent of diversification is positively related to profit rates. The estimated coefficients using one measure of diversification, E''_{5i}, were consistently positive but without acceptable significance. The other two measures of diversification, E_{5i} and E'_{5i}, consistently produced coefficients with the wrong (negative) sign and with large t-ratios in the case of E_{5i}. These results were not altered by the inclusion of relative firm size, E_{3i}, as an additional independent variable.

III. A Postscript

Viewing the diversified firm as a mere aggregate of "specialized firms" risks serious oversimplification and the statistical assumptions required to make the formulation operational are strong. Still, we believe that the resulting treatment of heteroskedasticity is less *ad hoc* than treatments found elsewhere. Importantly, the interdependence among error components for firms operating in the same market is taken into account in a systematic manner. It is our impression that this possibility has received scant attention in the literature though it should have been obvious in light of price theory and specification error. We cannot judge on a priori grounds how important it is to account for such interdependency in estimation, but the sensitivity of our statistical results to changes in the *omega* ratio suggest that those who ignore it proceed at their own peril.

Though we expect our empirical findings to be of some interest, the dictates of science call for replication of our results for different spatial-temporal domains. In this regard the similarity of our findings to those of the *FTC* is at least reassuring. In future work we believe researchers will need venture beyond merely wondering about the possible sensitivity of statistical results to industry definitions. If the food sector is representative, placing heavy reliance on Census definitions may be causing much mischief in research. Moreover, refinement of specific hypotheses and fuller specification of equations appear to be quite critical. Witness the widely disparate results for the absolute size variable depending on what other variables are included in the analysis.

APPENDIX

TABLE 1—RECONSTRUCTED CENSUS INDUSTRY DEFINITIONS, FOOD PROCESSING SECTOR

Animal feed, manufactured for commercial animals; Baby food in cans and jars; Baked goods, dry, crackers; Baked goods, dry, cookies; Baked goods, soft; Baking powder and yeast; Beans, canned, dry; Brewing; Chips, potato, corn and miscellaneous flour products; Confections, bars and specialties; Confections, bulk and packaged candy; Cereal, ready to eat; Cereal, to be cooked; Chocolate manufacturing; Coffee; Distilling, raw grain spirits; Distilling, beverages, blended and bottled; Dessert mixes; Fish and seafood, canned finfish; Fish and seafood, fresh and frozen finfish; Fish and seafood, fresh, frozen, and canned shellfish; Flavorings and extracts; Flavorings for soft drink bottlers; Fruit, canned and frozen; Fruit, dried; Grain mill products, wheat flour; Grain mill products, mixes and prepared flours; Grain mill products, refrigerated doughs; Gum, chewing; Juice, noncarbonated fruit and vegetable; Meat and poultry, fresh and processed; Meat sauces, all types, excluding catsup; Milk, canned evaporated and condensed; Milk, dried instant; Milk, fluid and dairy products; Margarine and butter; Nuts, shelled and salted; Oils, crude edible; Oils, refined edible; Peanut butter; Pet food; Pickles; Protein meals; Salad dressings, prepared; Soup, all forms; Spices; Sweeteners for human consumption; Tobacco, cigars; Tobacco, cigarettes; Tomato products, catsup; Tomato products, for cooking; Vegetables, canned and frozen; Vinegar; Wet corn milling.

TABLE 2—STANDARD INDUSTRIAL CLASSIFICATION OF FOUR- OR FIVE-DIGIT GROUPS USED FOR CENSUS FOUR-FIVE DIGIT INDUSTRY DEFINITIONS

2011; 2013; 2015; 2021; 20231; 20232; 2024; 20262; 2031; 20322; 20323; 20331; 20332; 20334; 20335; 20336; 20338; 20341; 20352; 20353; 20354; 2036; 20371; 20372; 20411; 20415+20455,[1] 2042, 20423; 20430; 20460; 2051; 20521; 20522; 20620+20630; 20711; 20712; 20713; 20714; 20716; 20720; 20722; 20730; 20820; 20851; 20853; 20860; 20871; 20873; 2091+2092+2093; 2095; 2096; 20961; 20962; 20991; 20992; 20993; 20994; 20996; 20999; 2111; 2121.

[1] See U.S. Senate Subcommittee on Antitrust and Monopoly, page 245.

REFERENCES

J. S. Bain, *Industrial Organization,* 2d ed., New York 1968.

W. J. Baumol, *Business Behavior, Value and Growth,* New York 1967.

R. Caves, *American Industry: Structure, Conduct, Performance,* 2d ed., Englewood Cliffs 1967.

W. S. Comanor and T. A. Wilson, "Advertising, Market Structure and Performance," *Rev. Econ. Statist.,* Nov. 1967, *49,* 432–40.

C. D. Edwards, "Conglomerate Bigness As a Source of Power," *Business Concentration and Price Policy,* Princeton 1955, pp. 331–52.

A. S. Goldberger, *Econometric Theory,* New York 1964.

M. Hall and L. Weiss, "Firm Size and Profitability," *Rev. Econ. Statist.,* Aug. 1967, *49,* 319–31.

J. B. Imel, "Structure-Profit Relationships in the Food Processing Sector," unpublished doctoral dissertation, Univ. Wisconsin 1971.

J. Johnston, *Econometric Methods,* New York 1963.

D. R. Kamerschen, "The Influence of Ownership and Control on Profit Rates," *Amer. Econ. Rev.,* June 1968, *58,* 432–47.

C. Kaysen and D. F. Turner, *Antitrust Policy: An Economic and Legal Analysis,* Cambridge, Mass. 1959.

A. C. Manchester, *The Structure of Fluid Milk Markets: Two Decades of Change,* U.S. Department of Agriculture, Economic Research Service, Report No. 137, Washington 1968.

F. Modigliani, "New Developments on the Oligopoly Front," *J. Polit. Econ.,* June 1958, *66,* 215–32.

G. J. Stigler, *Capital and Rates of Return in Manufacturing Industries,* Princeton 1963.

O. E. Williamson, "Selling Costs as Barrier to Entry," *Quart. J. Econ.,* Feb. 1963, *77,* 112–28.

Bureau of Advertising of American Newspaper Publishers Association, *Expenditures of National Advertisers in Newspaper Year—1967, Year—1960, and Year—1956,* New York 1967.

Federal Trade Commission, *Economic Report on the Influence of Market Structure on the Profit Performance of Food Manufacturing Companies,* Washington Sept. 1969.

Institute of Outdoor Advertising, Inc., *Outdoor Advertising Expenditures,* Norwalk, Conn. 1967.

Leading National Advertisers, Inc., *National Advertising Investments,* Norwalk, Conn. 1967; and 1960, 1956.

Moody's Industrial Manual, New York 1959; and 1960–68.

National Commission on Food Marketing, *Organization and Competition in the Milling and Baking Industry,* tech. study 5, Washington 1966.

———, *The Structure of Food Manufacturing,* tech. study 8, Washington 1966.

National Science Foundation, *Funds for Research and Development in Industry, 1959,* Washington 1962.

Radio Expenditure Reports, *Radio: Network and Spot Radio Advertisers: Estimated Expenditures—National and Regional Advertisers,* Larchmont, N.Y., 1960, 1967.

Television Bureau of Advertising, Inc., *Tenth Annual Spot Television Expenditure Report,* New York 1965; Also *Fifth Annual Report,* 1960 and *First Annual Report,* 1956.

U.S. Department of Commerce, Bureau of the Census, *1963 Census of Manufacturers, Vol. II: Industry Statistics, Part I: Major Groups 20–28,* Washington 1966.

U.S. Patent Office, *Index of Patents Issued from the United States Patent Office: 1958,* Washington 1959; also *1951–68.*

U.S. Senate Subcommittee on Antitrust and Monopoly, *Concentration Ratios in Manufacturing Industry 1963,* Washington 1966.

The Pareto Optimal Approach to Income Redistribution: A Fiscal Application

By GEORGE M. VON FURSTENBERG AND DENNIS C. MUELLER*

In the last few years several attempts have been made to reintroduce the income distribution question into the mainstream of economic literature. The most important of these were made by Harold Hochman and James Rodgers (1969, 1970), their commentators, Paul Meyer and J. J. Shipley, Richard Musgrave, and Robert Goldfarb, and more ambitiously, Mark Pauly (1970b). Treating income redistribution as a public good, these authors employ the corresponding established theory, based on consumer sovereignty and individual pursuit of gain, to determine a Pareto optimal level of redistribution. They thus dispense both with externally contrived social welfare functions and with theories of conflict and strategy in order to remove the dichotomy between allocation and distribution decisions from the Public Finance literature.

Though it may not be coextensive with the political issue,[1] the Pareto optimal approach to redistribution appears promising at first glance. If a significant amount of redistribution can be provided without anyone losing welfare, social conflict and the need for coercion can be minimized. However, before committing ourselves, the theoretical and practical viability of the Pareto optimal construction of the problem must be tested by working through to a fiscal application. In the next two sections of this paper we will therefore pay close attention to the assumptions which have to be made before a specific tax-transfer structure can be derived. To obtain such an operational solution, a utility function will be specified which, while arbitrary and thus purely exploratory in detail, nevertheless isolates the degree of altruism and other essential postulates required to justify any given transfer level within this framework. After demonstrating the basic method with a case involving three persons two of whom are transferors, a compulsory, though formally Pareto optimal, redistributive tax scheme is simulated in Section II for the *U.S.* economy as a whole.

The conditions for voluntary income transfers between two persons, outlined by Hochman and Rodgers, need not be recapitulated here. Uncoordinated or bilateral transfers are forthcoming if, and only if, the marginal utility of transferring a dollar of one's own income exceeds the marginal utility of retaining it. To allow for the possibility of altruism, the total incomes of the poor (Hochman-Rodgers) or a selected range of externality-generating commodities consumed by the poor (Pauly) are included as arguments in the preference functions of potential donors. In the two-person case with one rich and one poor, an optimal outcome can still be achieved without the aid of government organization. However, as Musgrave and Goldfarb have pointed out, unless there is a collective plan in which many are required to share in the cost of providing additional dollars to the poor, the amount of redistribution undertaken in society at large remains suboptimal. Moreover, *voluntary* burden-sharing agreements are im-

---

* Associate professors of economics at Indiana and Cornell University, respectively. We wish to thank Robert Tollison and an anonymous referee for helpful comments.

[1] See Musgrave.

possible to achieve among numerous potential transferors, since each person regards the amount of redistribution provided as independent of his particular contribution to this public good. Hence, if we are to pursue the justification and feasibility of government-enforced Pareto optimal redistribution, it is clearly necessary to go beyond the two-person case.

I. The Range of Pareto Optimal Solutions for Three Persons

As one approaches the fiscal setting, further stipulations have to be made. Taxes are assumed to be levied on the basis of individual incomes regardless of source, so that—distinctions by age and family-size aside—a single schedule applies to the same concept of income as is used to define the demand for public goods and for making transfers. This implies that only the progressivity of the schedule and the level of taxes and services can be varied through fiscal management. Applied to redistributive taxes, consumer sovereignty can thus be respected only if summary statements about the income elasticity of demand for making transfers are possible.[2] In addition, the progression of marginal rates of substitution between own and other income has to be sufficiently smooth[3] across the different income classes to yield tax structures which add up to a feasible political program.[4]

The knowledge that units with the same income will be taxed equally prevents dissimulation of preferences. At the same time, if the Pareto restriction is taken literally, genuine differences in tastes *within* income classes must be either assumed away or "averaged out." Given that preferences are homogeneous within each income group, one must still require independence of preferences between different groups of transferors and the absence of endogenous change in preferences, if the transfer decision is to be deduced from predetermined individual tastes.

If the previous conditions are met,[5] introduction of an ordinal utility function containing only income variables involves no further compromises or commitments. To clarify the process of public choice, one might first specify the kinds of utility functions implicit in political programs of the day. These policy functions could then be contrasted with the criterion function which is to be inferred from individual preferences. Although no such function has as yet been proposed, its outlines might conceivably be inferred from novel consumer surveys or more indirect evidence.[6] If the criterion function, with respect to which Pareto-preferred moves are defined, can be matched with one of the policy functions, the implied redistributive program would carry normative sanction.

To explore the potential of this approach, we choose the ordinal utility function for income group i,

---

[2] See David Winch, especially pp. 495–97.

[3] Koichi Mera uses a smoothness criterion to identify the sacrifice principle of taxation which is most compatible with the effective *U.S.* tax structure.

[4] An acceptable redistributive tax structure could not be found, for instance, if people cared mainly about relatives who tend to be members of the same group, or if charisma ("the charwoman who knits a shawl for the princess") is significant. See William Vickrey, pp. 40–44. More generally, if making transfers becomes a strongly inferior good beyond some high level of income, all conceivable Pareto optimal programs might outrage transferors with incomes below this level.

[5] In Section III we explore some of the implications of their not being met.

[6] Professor Robert W. Kilpatrick of Cornell University informs us that he has analyzed a time-series of responses to the Gallup question, asked since 1947: "What is the smallest amount of money a family of four (husband, wife, and two children) needs each week to get along in this community?" Preliminary indications are that the presumed poverty line rises systematically, but not strongly, with the income of the respondent. This would seem to imply that people take a slightly more generous view of poverty, the farther they are away from it. The general implication that people consider their own income in relation to that of lower-income families to determine Pareto optimal redistribution is built into the utility function specified below.

$$U_i' = Y_i^a \prod_{j=1}^{i} (Y_j/Y_i)^b,$$

to yield a minimum guaranteed income solution. While other income redistribution objectives can be envisaged, we believe that the guaranteed income objective is the one most in keeping with the voluntary transfer approach, and for that matter with current political debates. If the marginal utility to transferor i, of raising the income of j, is a decreasing function of Y_j, as our function implies, transfers will always be directed to whoever remains poorest. Of course, work incentive considerations might dictate otherwise. However, any deviation from the policy of bringing all those receiving transfers up to the same income level must be based on nondistributive criteria of efficiency and allocation which are beyond the scope of this analysis.

Beginning with three persons whose income is ranked from the lowest, such that $Y_1 < Y_2 < Y_3$, a tax scheme can be derived that makes the richer two share in the cost of financing the transfer to the lowest-income person. A range of Pareto optimal solutions results in a manner similar to the Johansen-Lindahl derivation.[7] As already explained, it is specified that the income of others, Y_j, is only taken into account by i, if it is less than his own income, Y_i. After taking logarithms the utility function becomes:

$$U_i = a \log Y_i + b \sum_{j=1}^{i-1} \log (Y_j/Y_i), \qquad Y_j < Y_i \tag{1}$$

Specifically, the utility functions for the three income classes are:

$$U_1 = a \log Y_1 \tag{2}$$

$$U_2 = (a - b) \log Y_2 + b \log Y_1 \tag{3}$$

$$U_3 = (a - 2b) \log Y_3 + b \log Y_1 + b \log Y_2 \tag{4}$$

Assume persons 2 and 3 bear c percent and $(1-c)$ percent of the cost of transferring S_1 dollars to individual 1, respectively. If the initial income levels are $Y_1{}^*$, $Y_2{}^*$, and $Y_3{}^*$, then the post-transfer incomes are:

$$Y_1 = Y^*{}_1 + S_1 \quad \text{and } dY_1 = dS_1, \tag{5}$$

$$Y_2 = Y^*{}_2 - cS_1 \quad \text{and } dY_2 = -\,cdS_1, \tag{6}$$

$$Y_3 = Y^*{}_3 - (1 - c)S_1 \quad \text{and } dY_3 = -\,(1 - c)dS_1 \tag{7}$$

Pareto optimal redistribution can be achieved as long as, for individuals 2 and 3, the change in utility from transferring income is greater or equal to zero. The recipient (1) always stands to gain, even at the margin, since a is positive. Differentiating equations (3) and (4),[8] substituting for the Y and dY, and dividing through by dS_1, we therefore obtain:

$$dU_2/dS_1 = -\,[(a - b)/(Y^*{}_2 - cS_1)]c + b/(Y^*{}_1 + S_1), \tag{3a}$$

$$dU_3/dS_1 = -\,[(a - 2b)/(Y^*{}_3 - (1 - c)S_1)](1 - c) + b/(Y^*{}_1 + S_1) - [b/(Y^*{}_2 - cS_1)]c \tag{4a}$$

If the coefficients of the utility function, a and b, and the initial income levels, Y^*,

---

[7] See Leif Johansen. As shown by Vincent Tarascio, pp. 10–12, Pareto himself used subjective social welfare functions of individuals at one point, in which the presumed utility levels of others appear as arguments. The resulting differential equations, or "offer curves," could easily form the nucleus of a "Pareto optimal" welfare redistribution theory.

[8] From the total derivative of (4),

$$dU_3 = [(a - 2b)/Y_3]dY_3 + [b/Y_1]dY_1 + [b/Y_2]dY_2,$$

it can immediately be seen that the utility function indeed implies a minimum guaranteed income solution. As long as Y_2 in the denominator is greater than Y_1, the utility increment derived by individual 3 will always be highest per dollar transferred to individual 1. Hence, no transfers will go to any higher-income person until all those with lower income have been brought up to his level.

are given, one can solve for those combinations of c and S_1 which reduce (dU/dS_1) to zero in both (3a) and (4a). The resulting optimal transfer loci, or offer curves, O_2 and O_3, are shown in Figure 1. In the diagram, in which the cost share c is plotted against the transfer amount S_1, the slope of indifference curves, such as I_2 and I_3, must be zero where they intersect their respective offer curve. Individual 2's indifference curves slope down from any such point, since any deviation from the optimal S_1 must be compensated through a reduction in c. Conversely, since individual 3's cost share is $(1-c)$, his indifference curves must be sloping up in both directions from any point on O_3. The locus of all points of tangency between indifference curves I_2 and I_3 yields the contract curve CD. Moving up this curve, individual 3's welfare level rises, while that of individual 2 falls.

Figure 1 is drawn for a equal to .9 and b equal to .1, but it can easily be seen from equations (3a) and (4a) that the location of the offer curves in the transfer diagram depends only on the ratio b/a. Increasing the relative altruism coefficient, b/a, shifts the offer curves outward since the welfare of transferors has become more responsive to raising the income levels of transferees. In the next section, b/a is allowed to vary to reveal the degree of altruism that would have to be present if different minimum guaranteed incomes could be provided in the United States by redistributing solely in Pareto optimal fashion. Conversely, given b/a and a stable political consensus supporting it, exogenous changes in the absolute and relative income distribution produced by technological progress or other sources of economic growth would have predictable effects on the desired transfer level.

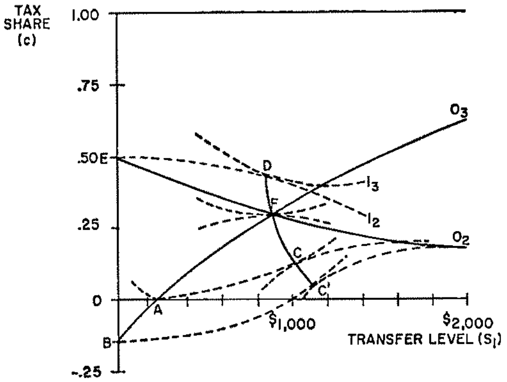


FIGURE 1. DERIVATION OF THE CONTRACT CURVE FOR REDISTRIBUTION FINANCED BY INDIVIDUALS 2 AND 3 IN THE RATIO c TO $(1-c)$

To complete the demonstration of the basic method used to derive explicit transfers, b/a is held fixed at its original level of 11.11 percent for the remainder of this section. If the initial income levels, Y_1^*, Y_2^*, and Y_3^*, are arbitrarily chosen to be \$2,000, \$8,000, and \$16,000, respectively, the intercept OB in Figure 1 is found to be $-(1/7)$, and OA equals \$250. The intersection of the offer curves occurs at point F, indicating a transfer level of \$889, one-third of which is paid by individual 2 and the other two-thirds by individual 3. Hence, the taxes of the two transferors are proportional to their income.

In this example, individual 2 would wish to transfer nothing at all, if he had to bear the whole burden so that c equals one. By contrast, individual 3 is willing to transfer a small amount OA when c equals zero.[9] As a matter of fact, individual 3 would even be willing to pay transaction costs[10] equal to $100B$ percent on top of the dollar transferred, if he would otherwise be prevented from making any transfers at all. Under those conditions, the contract curve would extend down to C'. In either case, a whole range of transfer levels and cost shares, with individual 2 paying from a little less than 50 percent to almost nothing, is seen to satisfy the requirement of Pareto optimality when a redistributive tax-transfer scheme is to be introduced. The Lindahl solution F, given by the intersection of the offer curves, has no more normative significance than any other point on the contract curve.

---

[9] Since we have here returned to bilateral distribution, with one person paying all, A might be called a Hochman-Rodgers transfer point, although it lies outside the range of Pareto optimal solutions in the three-person case. By varying the initial income level of one transferor, shifts in A could be produced which could be related to the original change in income to give precision to the Hochman-Rodgers concept of the transfer elasticity.

[10] If these costs take the form of a bribe paid to individual 2, the highest-income person, 3, derives some satisfaction also from adding to the income of individual 2. As a result, the negative intercept of O_3 would increase to 1/5 from 1/7, shown in Figure 1. Hence, conflict situations could be envisaged which would have to be resolved through payments to groups with unequal post-

Nonetheless, F does have a certain operational importance. For instance, if the size of the transfer is determined by polling individuals 2 and 3 about the levels each favors on the assumption that they pay c and $(1-c)$ percent of the transfer cost, respectively, then the only matching selection of amounts is given by F.[11] For higher values of c, individual 2 favors a lower transfer level, individual 3 wants a higher one, and the actual transfer sum provided would have to be worked out through bargaining among the two transferors. A segment of the contract locus between the offer curves would indicate the optimal results of compromise. For lower values of c the reverse holds and the same scope for game strategies exists. Hence, unanimous agreement on transfer levels can be obtained only if compulsory taxes are imposed at the cost shares given by F. Because of preference registration problems, the location of F would be harder to determine in the large number case; but setting the corresponding obligatory tax shares would still produce unanimity on amounts.[12] Furthermore, F is the only point for which numerical solutions can be obtained in the calculations that follow. Tax shares derived from the offer curves sum to unity only at the transfer level associated with F.

II. An Application to the *U.S.* Income Distribution

Before the previous model can be applied to the *U.S.* economy, a realistic condensation of the American income distribution must be obtained. Our aim is to delineate a manageable number of income classes of varying width but equal weight so as to include the same percentage of families in each. We can then take the estimate of mean income in each class, derive the offer curves, and solve for a point analogous to F.

Instead of arbitrarily fixing the level of a and b in the utility function, which ought to be based on (unknown) individual preferences, we substitute for S by imposing minimum guaranteed income levels of current political interest. We then inquire what the relation between a and b would have to be to imply this particular

---

transfer incomes and still not violate the Pareto criterion. In Figure 1, no such payments are actually visible, since C' is placed above the horizontal axis. Nevertheless, the relevance of the theoretical consideration is obvious in explaining the spill over into middle-income classes, which is observed in many government in-kind transfer programs with basically redistributive rationale.

[11] See Martin McGuire and Henry Aaron, pp. 32–33.

[12] If making transfers conferred benefits on donors because they took joy in, or heavenly rewards for, their own giving, the transfer activity would have all the characteristics of private goods. No benefits would accrue to transferors jointly from their own giving and the giving of others, and hence there would be no exclusion problem. In the model applied in this paper, transferors benefit whenever lower-income persons receive supplements; they are indifferent to the identity of the giver. With jointness thus admitted and exclusion not feasible, F can no longer be identified as a stable equilibrium, rather, non-core outcomes are likely. With a large number of taxpayers it is always advantageous for some taxpayer to form a coalition of one and refuse to pay taxes, since no rich person can be denied the benefits of seeing the poor's lot improved by income transfers from the other rich. However, if exclusion were feasible, jointness in consumption alone would not prevent F from being the politically relevant allocation in the core. On this, see Mark Pauly (1967, 1970a) and Duncan Foley.

solution. Since the equations for the offer curves are homogeneous of degree zero in a and b, only the ratio b/a and the tax cost shares of transferors c, enter as unknowns. Adding the condition that the tax shares add to unity, to the offer curves obtained for each class of transferors permits one to solve the system.

Charles Metcalf has shown how log-normality can be achieved through displacement of the *U.S.* income distribution. Dividing both the upper income bound of the first population decile and the lower bound of the highest decile by median family income, he found values of .31 and 2.00, respectively, to be representative for the postwar period. In 1970, the median income of primary families is expected to be around $10,000 so that the corresponding decile boundaries would be $3,100 and $20,000. Under these conditions, a displacement factor of $12,258 must be added to obtain a lognormal distribution, judged only by the two decile measures. The mean of the displaced distribution is $22,258, while $15,358 and $32,258 hold at the respective decile boundaries. Since *log* (22,258/15,358), or, symmetrically, *log* (32,258/22,258), equals 1.28 times the standard deviation of the logarithms of the displaced incomes values, if 40 percent of families fall into each range, the standard deviation can be calculated. Choosing income classes the width of a decile, class boundaries and means are readily obtained using z-values of specified areas under the normal curve $[z=(x-\mu)/\sigma]$. After subtracting the displacement factor, the resulting income levels are shown in Table 1.

We can then characterize the *U.S.* income distribution as if it consisted of ten individuals and derive offer curves as in Section I. Using the symbol c_i for the tax cost share at the mean income level of the ith population decile, offer curves are specified for either eight or nine transferors, depending on whether the income floor to be provided is above or below $4,230, the mean of the second decile. The difference between the guaranteed minimum and the mean income of $1,568 in the first decile yields the transfer amount S. For example, assuming a floor of $3,000, or S of $1,432, and setting a equal to any level, such as .9, while leaving b variable gives a triangular expansion of equations, comparable to (3a) and (4a). These are equations (8a), (8b), . . . , (8i), and (9).

With the last restriction, stating that the tax shares sum to unity, the ten non-linear equations above can be solved for the nine tax shares and b by computer iterations. The results are shown in Table 2 for income floors ranging from $1,600 to $5,000. It is noted that the distribution of costs between transferors is independent

$$(8a)\qquad 0 = -c_2(.9 - b)/(4{,}230 - 1{,}432c_2) + b/3{,}000$$

$$(8b)\qquad 0 = -c_3(.9 - 2b)/(6{,}052 - 1{,}432c_3) + b/3{,}000 - c_2b/(4{,}230 - 1{,}432c_2)$$

. . . .

$$\begin{aligned}(8i)\qquad 0 = &-c_{10}(.9 - 9b)/(23{,}578 - 1{,}432c_{10}) + b/3{,}000 - c_2b/(4{,}230 - 1{,}432c_2)\\ &- c_3b/(6{,}052 - 1{,}432c_3) - c_4b/(7{,}652 - 1{,}432c_4) - c_5b/(9{,}205 - 1{,}432c_5)\\ &- c_6b/(10{,}825 - 1{,}432c_6) - c_7b/(12{,}628 - 1{,}432c_7) - c_8b/(14{,}802 - 1{,}432c_8)\\ &- c_9b/(17{,}791 - 1{,}432c_9)\end{aligned}$$

$$(9)\qquad \sum_{i=2}^{10} c_i = 1$$

TABLE 1—A REPRESENTATIVE INCOME DISTRIBUTION FOR FAMILIES, ESTIMATED IN 1970 DOLLARS[a]

Decile	Range	Class Median[b]
1	$0- 3,100	$1,568
2	3,101- 5,187	4,230
3	5,188- 6,865	6,052
4	6,866- 8,427	7,652
5	8,428-10,000	9,205
6	10,001-11,695	10,825
7	11,696-13,650	12,628
8	13,651-16,142	14,802
9	16,143-20,000	17,791
10	20,001- ∞	23,578

[a] For the period 1949-65, Metcalf found that the annual income level at the bottom decile cut-off amounts to 31 percent of the median income of families with male head and wife not in the labor force. Such families account for almost 60 percent of all families. Similarly, the income level determined such that 90 percent receive less came to 200 percent of median family income on the average. Setting the latter equal to $10,000 to approximate 1970 levels then yields an income distribution whose structure is representative for the postwar period.

[b] The class medians, which are used as estimates of the class means, imply an average family income of $10,834, compared to $10,953 expected for the displaced lognormal distribution with variance .0838 in natural logarithms.

of the level of the income floor as long as only families in the lowest decile of the income distribution (represented by a single person) receive transfers. If the minimum guaranteed income level rises above $4,230 so that families in the second decile turn into transferees, new cost shares are established which would again remain constant until the floor rises above $6,052, the mean of the third decile. Furthermore, the cost shares imply that a proportional tax rate, shown on the last line of Table 2, should be imposed on all transferors. This consequence is inevitable if the utility functions are of the generalized Cobb-Douglas type, as can be seen by inspecting the original equations of the offer curves.[13]

Because of the loss of detail involved in representing the *U.S.* income distribution by a single measure per decile, estimates of the total federal budget cost of providing any given income floor cannot normally be derived from our results. For instance, if the floor happened to be $1,568, our specification would imply that no transfers are needed at all although the relative altruism coefficient, b/a, still amounts to over 1.4 percent. In fact, around 5 percent of all families will have less than $1,568, the mean (median) of the first decile. Hence, a meaningful estimate can be made only at the upper bound of the first decile, or at an income level of $3,100.[14]

Multiplying the difference between $3,100 and $1,568 by one-tenth of the number of families in the United States, or 5.2 million, yields a budget cost estimate of almost precisely $8 billion for 1970. In this, around 12 million unrelated individuals are excluded and all problems arising from the definition of income or from differences in the size and age structure of families are disregarded. Furthermore, the coefficient b would have to amount to exactly 2.86 percent of a in the criterion function, if $3,100 were to be the income floor desired by transferors acting in concord. It should be noted that this assumes a fairly sizable degree of altruism on the part of all those with incomes above $3,100, for it implies that they are willing to transfer 1.38 percent of their income to those in the bottom decile.

III. Evaluation

Having shown what is necessary to implement the Pareto approach to redistribution, we close with an examination of some of the basic assumptions on which it rests and consider conceptual and practical problems to which these assumptions give rise.

---

[13] The proof is based on showing that if the income of one group of transferors is k times as large as that of another, their tax share must also be k times as large to satisfy the equations.

[14] The sum of the absolute deviations of income from the upper bound of any distribution is equal to the distance between that limit and the mean, multiplied by the number of observations.

Table 2—Solutions of Hypothetically Pareto Optimal Tax-Transfer Programs for the United States

Income Floor	\$1,600	\$2,000	\$2,500	\$3,000	\$3,500	\$4,000	\$4,500	\$5,000
b/a	1.48%	1.84%	2.31%	2.77%	3.23%	3.69%	4.16%	4.61%
Tax Shares[a]								
c_2	.0396						—	
c_3	.0567						.0590	
c_4	.0717						.0746	
c_5	.0862						.0898	
c_6	.1014						.1056	
c_7	.1183						.1231	
c_8	.1386						.1444	
c_9	.1666						.1735	
c_{10}	.2209						.2300	
Implied Tax Rate	.03%	.40%	.87%	1.34%	1.81%	2.28%	3.12%	4.10%

[a] The tax shares sum to unity. At an income floor of \$3,000, they imply, for instance, that families in the second decile ($Y=\$4,230$) pay \$56.71 (3.96% of \$1,432) in taxes, while families in the tenth decile with income of \$23,578 pay \$316.33 (22.09% of \$1,432). In either case, taxes amount to 1.34% of income. Income inequalities are never reversed as a result of the transfers.

First there is the assumption that individual evaluations of need are based on the income levels of the poor and that redistribution takes the form of income transfers. Presumably, what bothers the rich is not that the money incomes of the poor (including annuitized net worth, if any) are low, but that, largely because of this, their realized welfare is thought to be low. Already at the outset, it would seem difficult to get transferors to agree on what cash-equivalent income, excluding taxes but including the value of all benefits derived from government, is to be regarded as affording the same level of welfare, for instance, to a public housing tenant in New York and to a tenant farmer in Mississippi. Then there will be debate over how any desired improvement can best be achieved. Income maintenance can be provided either through continual income transfers,[15] once and for all asset transfers, or increased socialization of ownership and services. Alternatively, job training programs and educational reforms can be used to increase earnings capacity without immediately affecting measured income. Therefore, even if all the rich agree on the desirability of redistribution in some form, the Pareto criterion might have to be violated if no agreement can be reached on the underlying welfare calculus and the appropriate means for carrying out such a program.

Another obstacle to the imposition of a Pareto optimal redistributive tax arises because of the practical need to tax all members with equal (taxable) incomes at the same rate and to preserve smoothness in the tax structure vertically. This problem exists, of course, with every public good. Among individuals with diverse tastes, some can always be found whose welfare is reduced by the combined tax package and public goods program they are forced to take. The problem is conceptually more serious in the case of a redistributive tax, however.

First, if one simply ignores the losers, one implicitly assumes that the benefits to those higher-income persons favoring the measure are sufficient to outweigh the losses of the others. Interpersonal utility comparisons are obviously involved here and one reverts to the old elitist approach to redistribution—where the elite is now a simple majority of all potential transferors. Once the possibility of this kind of conflict is admitted, it becomes logically difficult to exclude potential transferees from joining the altruistic elite or from forming a power block of their own. We are then thrust back to a situation of class conflict over the level of redistribution, where one group simply takes income from another (see Musgrave).

---

[15] For Pareto optimal programs involving price reductions on selected commodities consumed by the poor, see Pauly (1970b).

Second, one could argue that although the amount of any one public good supplied may not be Pareto optimal, the provision of all public goods is. Voters can trade away votes on issues they consider less important, increase their probability of winning on the remaining issues, and thus raise the likelihood that they will experience a net gain from the total package of public goods supplied (see Buchanan and Tullock, Coleman, and Mueller). Consistent with Hochman and Rodgers' shifting of redistribution into Musgrave's allocation branch, one might thus hold that differences in the preferences of donors in any income class can be ironed out by grouping redistribution with other public goods issues and allowing vote trading.

However, this is not a satisfactory solution if the primary function of redistribution is to improve the welfare position of the poor not just in isolation, but relative to the rich. If the opponents to redistribution are compensated by allowing them a greater influence on other public goods through vote trading, they may be left with a relatively greater command over private and public goods after redistribution—and vote trading—than they had before.[16] Thus, if the goal of redistribution is a narrowing of differences in welfare levels realized by rich and poor, the redistribution issue must be separated from the class of allocation issues. Otherwise the preceding goal might be compromised in the attempt to preserve Pareto optimality through logrolling.[17]

In summary, a meaningful program of redistribution, with the welfare of the poor improving relative to that of the rich, can only be achieved through a Pareto optimal tax-transfer program if all transferors approve of the amount and form in which redistribution should be carried out at given tax shares. Barring this, redistribution always involves the imposition of a program by some elite, if it is to be effective. Conversely, strict enforcement of the Pareto criterion under such conditions results in preservation of the *status quo ante*. With a less than universal desire for making transfers, the main objective tends to be thwarted as potential losers must be compensated via logrolling.[18]

---

[16] Redistributive tax measures will be passed along with expenditure bills that favor the rich. Since the package deal may confer uneven "gains from trade," it may still be possible to improve the relative welfare of the poor, for instance, by keeping the welfare level of transferors unchanged. In actuality, however, with the rich disagreeing on the desirability of transfers, relative bargaining strengths and skills among the rich and between rich and poor will determine the division of the gains. Thus, there can be no presumption that the rich will agree to the deal if they do not share its net benefits.

[17] If there is no need for logrolling because preferences for redistribution are perfectly homogeneous within each group of transferors, the relative welfare of the poor will always be improved at least at the margin of the transfer when donors are indifferent but recipients still unsated. However, the slight equalization of relative welfare levels which may occur under conditions most favorable to Pareto optimal redistribution is entirely incidental to the approach. Global concepts of social justice, applied to the whole income distribution, violate the rules of the game if they lead to coercion of transferors either by other transferors or by the class of recipients.

[18] In spite of these difficulties, the attempt to capitalize on altruism may be helpful in shaping positions on matters of social policy. For instance, those who favor redistribution for purely selfish reasons can benefit in the knowledge that not all transferors are hurt since some are in favor of this program. Once the existence of income-systematic altruism has been identified, potential transfer recipients should take care not to bargain away all the gains from a redistributive measure in an effort to get it passed.

REFERENCES

J. M. Buchanan and G. Tullock, *The Calculus of Consent,* Ann Arbor 1962.

J. S. Coleman, "The Possibility of a Social Welfare Function," *Amer. Econ. Rev.,* Dec. 1966, *56,* 1105–22.

D. K. Foley, "Lindahl's Solution and the Core of an Economy with Public Goods," *Econometrica,* Jan. 1970, *38,* 66–72.

R. S. Goldfarb, "Pareto Optimal Redistribution: Comment," *Amer. Econ. Rev.,* Dec. 1970, *60,* 994–96.

H. M. Hochman and J. D. Rodgers, "Pareto Optimal Redistribution," *Amer. Econ. Rev.,* Sept. 1969, *59,* 542–57.

——— and ———, "Pareto Optimal Redistribution: Reply," *Amer. Econ. Rev.,* Dec. 1970, *60,* 997–1002.

L. Johansen, "Some Notes on the Lindahl Theory of Determination of Public Expenditures," *Int. Econ. Rev.,* Sept. 1963, *4,* 346–58.

M. C. McGuire and H. Aaron, "Efficiency and Equity in the Optimal Supply of a Public Good," *Rev. Econ. Statist.,* Feb. 1969, *51,* 31–39.

K. Mera, "Experimental Determination of Relative Marginal Utilities," *Quart. J. Econ.,* Aug. 1969, *83,* 464–77.

C. E. Metcalf, "The Size Distribution of Personal Income During the Business Cycle," *Amer. Econ. Rev.,* Sept. 1969, *59,* 657–68.

P. A. Meyer and J. J. Shipley, "Pareto Optimal Redistribution: Comment," *Amer. Econ. Rev.,* Dec. 1970, *60,* 991–93.

D. C. Mueller, "The Possibility of a Social Welfare Function: Comment," *Amer. Econ. Rev.,* Dec. 1967, *57,* 1304–11.

R. A. Musgrave, "Pareto Optimal Redistribution: Comment," *Amer. Econ. Rev.,* Dec. 1970, *60,* 991–93.

M. V. Pauly, "Clubs, Commonality, and the Core: An Integration of Game Theory and the Theory of Public Goods," *Economica,* Aug. 1967, *34,* 314–24.

———, (1970a) "Cores and Clubs," *Publ. Choice,* fall 1970, *9,* 53–65.

———, (1970b) "Efficiency in the Provision of Consumption Subsidies," *Kyklos,* 1970, *23,* Fasc. 1, 33–58.

V. J. Tarascio, "Paretian Welfare Theory: Some Neglected Aspects," *J. Polit. Econ.,* Jan./Feb. 1969, *77,* 1–20.

W. S. Vickrey, "One Economist's View of Philanthropy," in F. G. Dickinson, ed., *Philanthropy and Public Policy,* Nat. Bur. Econ. Res. Conference, New York 1962, 31–71.

D. M. Winch, "Pareto, Public Goods and Politics," *Can. J. Econ.,* Nov. 1969, *2,* 492–508.

Stability in Periodic Markets

By NORMAN P. OBST*

In the dynamic economic market models usually postulated, trading occurs exclusively continuously or trading occurs exclusively at separate, equally spaced points. This paper analyzes a market which is periodically open and closed (e.g., on evenings and weekends), such as the stock market. Trading occurs continuously when the market is open; it is prohibited when the market is closed. The existence of groups of traders who operate on information obtained only when the market is open, and of different groups of traders who operate on information which can be obtained only after the market closes, implies that the exclusive use of either differential or difference equations will result in an inadequate representation of the dynamic behavior of prices. In what follows, stability of the periodic market described above will be studied by using a mixed difference-differential equation model.

Four separate groups of traders are postulated. The first two groups make their decisions overnight and affect the opening price of the next day. The second two groups trade during the remainder of the day. The first and third groups of traders represent "investors" who consider current prices only. Their demand functions are assumed to be linear in the price at which they trade. The second and fourth groups of traders represent either emotional investors or "speculators" who consider only past prices in their decisions.[1] Their demand functions consist of an expectations term representing past price behavior. The second group is assumed to be influenced by the change between the previous two closing prices (last price traded during a day). The fourth group is assumed to be influenced by the rate of change of prices during the day. The primary advantage of the particular forms of these assumptions is the simplicity of the mathematics.

The market mechanism is assumed to be a zero excess demand process defined on a two-good exchange market, with one of these goods the numeraire. The demand functions of each participant are sent to a central agent (specialist) who sets the market price where excess demand is zero. The demand functions of the first two groups determine the opening price each day from the closing prices of the previous two days. Given the opening price, the prices during the remaining part of the day are obtained by setting the total demand of groups three and four equal to zero. This system generates a closing price which, in turn, determines the opening price for the next day. The equilibrium price of this type of dynamic mechanism is a price at which the system will remain if that price is generated for a sufficiently long time.

Stability of the discontinuous model will be defined in terms of the equilibrium closing price. It will be shown that if the closing prices each day converge to their equilibrium value then the opening prices

---

* Assistant professor of economics at the University of Washington, Seattle. I would like to thank Stanley Reiter for his invaluable guidance. Special thanks also to Edward Ames and to an anonymous referee for many helpful suggestions.

[1] The question of profitability of these traders is not relevant. One can imagine a steady stream of unprofitable emotional investors with new ones constantly replacing those running out of assets.

each day converge to their equilibrium value. It will also be expedient to consider the stability of the associated continuous model. The discontinuous model will be said to possess intraday stability if the system in which the closing price each day is the same as the opening price the next day is a stable system. Alternatively, intraday stability can be thought of as the stability of a system in which intervals of time between days are eliminated so that the market never closes and time is treated continuously. One of the important results obtained from analysis is: If the discontinuous system is intraday stable, then the values of the parameters of that system will be closer to those which lead to stability of the entire model, the larger the fraction of a day's time devoted to trading. Alternatively, if the system is intraday unstable, then the longer the time devoted to trading, the further away will be the values of the parameters from those which lead to overall stability. Thus it is possible to create an unstable system from a stable one (or the reverse) by lengthening or shortening the time devoted to trading during the day.

In what follows a discrete model, a continuous model, and the discontinuous model discussed above will be presented. The same principles will be used in discussing a weekly model. Conditions for stability will be derived.

I. A Discrete Model

In this model it will be assumed that there are two groups in the market, one of which reacts to past price behavior;

$$D_{1t} = -aP_t + b$$

$$D_{2t} = c(P_{t-1} - P_{t-2})$$

where D_{it} is the quantity demanded by group i at time t; $i=1, 2$; $t=1, 2, 3, \ldots,$; $a>0$, $b>0$, and $c\neq 0$.

Extrapolative expectations have been introduced: $c>0$ means previous price change leads people to believe prices are going to change in the same direction in the future; $c<0$ means people believe prices will move in the reverse direction, and expectations may be said to be conservative.[2]

Price is set by a central agent so that the total quantity demanded at each point in time is zero. Setting the sum of D_{1t} and D_{2t} equal to zero, we get a solvable difference equation of the form

$$P_t + a_1P_{t-1} + a_2P_{t-2} = G \tag{1}$$

where $a_1 \equiv -c/a$, $a_2 \equiv c/a$ and $G \equiv b/a$.

The necessary and sufficient conditions for stability of equation (1) are

$$\begin{aligned} &\text{(i) } 1 + a_1 + a_2 > 0, \\ &\text{(ii) } 1 - a_1 + a_2 > 0, \\ &\text{(iii) } 1 - a_2 > 0 \end{aligned} \tag{2}$$

The path of prices in the discrete model converges to its equilibrium price if, and only if, $-1/2 < c/a < 1$.

II. A Continuous Model

In this model, t is defined on the set of all positive real numbers. Again there are two groups:

$$D_3'(t) = -a'P(t) + b'$$

$$D_4'(t) = c'\frac{dP}{dt}$$

where $D_i'(t)$ is the quantity demanded by group i at time t; $i=3, 4$; $a'>0$, $b'>0$, and $c'\neq 0$.

By setting the sum of $D_3'(t)$ and $D_4'(t)$ equal to zero, we get a solvable differential equation. Price converges to its equilibrium value if, and only if, $c'<0$.

III. A Discontinuous Model

The following notation will be necessary:

Let R be the set of real numbers.

---

[2] See the article by Alain C. Enthoven and Kenneth J. Arrow.

Let $N \equiv \{1, 2, 3, \ldots,\}$.

Let *day* $t \equiv [t-1, t]$, $t \in N$ where $[t-1, t] \equiv \{u \in R \mid t-1 \leqq u \leqq t\}$.

Let *half-open day* $t \equiv (t-1, t]$, $t \in N$; that is, day t excluding opening time $t-1$; $(t-1, t] \equiv \{u \in R \mid t-1 < u \leqq t\}$.

Let $P(v)$ be the price at time v, $v \in$ half-open day t for some $t \in N$.

Let $P(t) \equiv P_t$, $t \in N$, and be called the closing price on day t.

Let P_t^* be the price at opening time $t-1$ (opening price) of day t, $t \in N$.

This model will represent continuous trading periods separated by nontrading intervals. It is clarifying to consider a market with trading occurring during a fixed period (say five hours), and with no trades permitted between the close of the market and its next day's opening.

The existence of four groups in the market is assumed: two form their decisions between days, and buy or sell at the opening price of a day; two buy and sell during the half-open day.

Let D_{it}^* be the quantity demanded by group i at time $t-1$ of day t, $t \in N$, $t > 2$; $i = 1, 2$.

Let $D_i(v)$ be the quantity demanded by group i at $v \in$ half-open day t, $t \in N$, $t > 2$; $i = 3, 4$.

Interday

$$(3) \qquad \begin{aligned} D_{1t}^* &= -aP_t^* + b \\ D_{2t}^* &= c(P_{t-1} - P_{t-2}) \end{aligned}$$

Intraday

$$(4) \qquad \begin{aligned} D_3(v) &= -a'P(v) + b' \\ D_4(v) &= c' \frac{dP}{dv} \end{aligned}$$

For both sets of equations (3) and (4), the parameters a, b, a', and b' are positive and the parameters c and c' are non-zero.

Equation set (3) determines the opening price each day from the closing price of the previous two days. Equation set (4) determines the price during the half-open day (and hence the closing price) from the opening price determined by equation set (3). Since it is assumed that group 4 above takes the opening price of each day into consideration through the dP/dv term, the following assumption is made: Let $v \in$ half-open day t for $t \in N$; then

$$(5) \qquad \operatorname*{limit}_{v \to t-1} P(v) = P_t^*$$

The intraday equations determine a first-order differential equation. The solution path of prices for each half-open day is dependent upon one constant which is determined by a specification of the initial price for that day. Equation (5) specifies that initial price as the opening price of the day determined by the interday equations.

The discontinuous model will now be solved for a difference equation representing the path of closing prices each day.

Setting the sum of $D_3(v)$ and $D_4(v)$ equal to zero, and solving for $P(v)$ we get

$$(6) \qquad P(v) = l_t e^{a'v/c'} + \frac{b'}{a'}$$

where $v \in$ half-open day t, $t \in N$, and $t > 2$. By equation (5), l_t is determined so that

$$P_t^* = \lim_{v \to t-1} \left(l_t e^{a'v/c'} + \frac{b'}{a'} \right)$$

Therefore

$$(7) \qquad P_t^* = l_t e^{a'(t-1)/c'} + \frac{b'}{a'}$$

From equation (6) and the definition of P_t we have

$$(8) \qquad P_t = l_t e^{a't/c'} + \frac{b'}{a'}$$

Substituting l_t of equation (7) into equation (8) we obtain

$$P_t = \left(P_t^* - \frac{b'}{a'}\right) e^{a'/c'} + \frac{b'}{a'} \tag{9}$$

Equation (9) expresses the closing price each day in terms of the opening price for that day.

Setting the sum of D_{1t}^* and D_{2t}^* equal to zero we get

$$P_t^* = \frac{cP_{t-1}}{a} - \frac{cP_{t-2}}{a} + \frac{b}{a} \tag{10}$$

Substituting (10) into (9) results in

$$\begin{aligned} P_t &= \frac{c}{a} e^{a'/c'} P_{t-1} - \frac{c}{a} e^{a'/c'} P_{t-2} \\ &\quad + \left(\frac{b}{a} - \frac{b'}{a'}\right) e^{a'/c'} + \frac{b'}{a'} \end{aligned} \tag{11}$$

The discontinuous model will be called *stable* if, and only if, the closing prices converge to their equilibrium value. Substituting for P_t, P_{t-1}, and P_{t-2} in equation (11) using equation (9), it is easy to show that the discontinuous model is stable if, and only if, the opening prices for each day converge to their equilibrium.

Equation (11) represents the path of closing prices each day. Using equations (2) it can be demonstrated that this path, and hence the discontinuous model, is stable if, and only if,

$$-\frac{1}{2} < \frac{c}{a} e^{a'/c'} < 1$$

This result is interesting in that parameter values which would cause instability in a discrete model or in a continuous model may not necessarily cause instability in the discontinuous model. If $c' > 0$, for instance, the intraday equations (4) would give an unstable system if time were to be treated as exclusively continuous. But in the discontinuous model

$$-\frac{a}{2} e^{-a'/c'} < c < a e^{-a'/c'}$$

will guarantee stability. Similarly, if the exclusively discrete model is unstable there may be values of c' which lead to overall stability. The significance of this is that either intraday or interday behavior may compensate for the instability of the other.

IV. Stability of Discontinuous Models with Unequal Trading Lengths

In this section the discontinuous model will be generalized. The relative stability of models with different lengths of time devoted to trading each day will be studied. The changes in notation will appear first.

Let $0 \leqq x \leqq 1$ indicate the fraction of a unit of time devoted to trading.

Let *day* $t \equiv [t-1, x+t-1]$, $t \in N$.

Let *half-open day* $t \equiv (t-1, x+t-1]$, $t \in N$.

Let $P(x+t-1) \equiv P_t$, $t \in N$.

All other definitions and assumptions remain unchanged. It should be noticed that the definitions and notation presented earlier are a special case of those above; that is, when $x = 1$.

The solution of this system is obtained in the same manner as in the special case. Equation (8) becomes

$$P_t = l_t e^{a'(x+t-1)/c'} + \frac{b'}{a'}$$

Therefore equation (9) is now

$$P_t = \left(P_t^* - \frac{b'}{a'}\right) e^{a'x/c'} + \frac{b'}{a'} \tag{12}$$

Combining equations (10) and (12) we obtain

$$\begin{aligned} P_t &= \frac{c}{a} e^{a'x/c'} P_{t-1} - \frac{c}{a} e^{a'x/c'} P_{t-2} \\ &\quad + \left(\frac{b}{a} - \frac{b'}{a'}\right) e^{a'x/c'} + \frac{b'}{a'} \end{aligned} \tag{13}$$

The generalized version is stable if, and only if,

$$-\frac{1}{2} < \frac{c}{a} e^{a'x/c'} < 1 \tag{14}$$

This result implies that if c' is positive, then shortening the length of the day has a stabilizing effect; if c' is negative, then increasing the length of the day has a stabilizing effect. It has been remarked above that c' negative is a necessary and sufficient condition for the exclusively continuous (intraday) model to be stable. Therefore the above result may be interpreted as follows: If the intraday model is unstable, then shortening the length of the day has a stabilizing effect; if the intraday model is stable, then increasing the time devoted to trading each day has a stabilizing effect. Whether the stability of the model is increased by an amount sufficient to create a stable model from an unstable one depends on the values of c, c', a, and a'. The possibility remains, however, that by varying the length of time devoted to trading each day a stable system may become unstable or an unstable system may become stable.

It is useful to notice here that the discrete and continuous models presented above are special cases of the generalized discontinuous model. If x is set equal to zero, the opening price and the closing price each day are the same; the day consists of one point at which all trading occurs, hence trading is at discrete steps. The correct conditions for stability of the discrete model are obtained by letting x equal zero in (14). If x is allowed to become arbitrarily large, the length of the day becomes arbitrarily large. The closing price is attained only in the limit. Time is, in effect, exclusively continuous. The correct conditions for stability of the continuous model are obtained by letting x be arbitrarily large in (14).

V. A Weekly Model

It is apparent from the media that there is some interest in obtaining weekly prices indicating the change in closing price from the preceding week. In this model, a fifth and sixth group who are weekly traders will be assumed. Both will transact at the opening price of the week only. It will be assumed that group 5 trades the same quantity at the opening price of the week and has the same equilibrium price as group 1. Extrapolative expectations with regard to closing weekly prices are assumed for group 6. While these assumptions are restrictive, the results are more general.[3] The effect of changing the number of days in the trading week upon the stability of the weekly closing price path will be studied. The new notation will be presented first.

Let $n \in N$ represent the number of days in a week.

Let *week* $u \equiv \{$day $1-n+nu$, day $2-n+nu, \ldots,$ day $nu\}$, $u \in N$.

Let $\bar{P}_u$ be the closing price of week u.

Let P'_u be the opening price of week u.

Let D'_{iu} be the quantity demanded by group i at the opening price of week u, $u \in N$, $u > 2$; $i = 5, 6$.

Throughout this section on the weekly model it will be assumed that $x = 1$.

It should be clear from the above definitions that the opening price of week u is the opening price of day $1-n+nu$ and that the closing price of week u is the closing price of day nu. In notation this is written as

---

[3] It may be more intuitive to some to assume instead that group 2 traders react to weekly closing prices as well as daily closing prices at the opening price of the first day of the week. In that case, there would be no group 5. Group 6 demand would be replaced by an equivalent demand added to group 2's demand function at the opening price of the week. This case can be handled with no change in notation up to equation (31) except the value of α changes. The stability conditions and the corresponding results obtained are qualitatively the same although quantitatively different.

$$P'_u \equiv \overset{*}{P}_{1-n+nu} \quad \text{and} \quad \bar{P}_u \equiv P_{nu}$$

The weekly model is composed of the interday equations (3), the intraday equations (4), and the interweek equations below.

Interweek

$$(15) \qquad \begin{aligned} D'_{5u} &= -aP'_u + b \\ D'_{6u} &= c''(\bar{P}_{u-1} - \bar{P}_{u-2}) \end{aligned}$$

For equation set (15) the parameters a and b are positive as before and c'' is nonzero. All other parameters of the weekly model have the same restrictions as in the daily discontinuous model. The weekly model will now be solved for difference equations representing the path of closing prices each week.

There are two cases: the first will apply when $(c/a)e^{a'/c'}$ does not equal four. In that case the roots of the auxiliary equation

$$(16) \qquad m^2 - \frac{c}{a} e^{a'/c'} m + \frac{c}{a} e^{a'/c'} = 0$$

of the homogeneous version of the difference equation (11) are real and unequal or complex conjugates. Difference equation (11) is the solution to the daily (intraweek) discontinuous model. It is applicable to the weekly model for any day t except $t=1-n+nu$, $u \in N$ since equation (10) holds for every $t \neq 1-n+nu$, $u \in N$ and equation (9) always holds true.

For ease of analysis, weeks u, $u-1$, and $u-2$ will be written explicitly below as Definition A.

Let m_1 and m_2 be the roots of the auxiliary equation (16). Then the solution to the daily closing difference equation is

$$P_t = c_1 m_1{}^t + c_2 m_2{}^t$$

The constants c_1 and c_2 are chosen so the equation yields the correct values for P_Φ and $P_{\Phi+1}$, $\Phi \in N$. The solution can be written as

$$(17) \qquad \begin{aligned} P_{\Phi+t} = {} & \left(\frac{P_{\Phi+1} - P_\Phi m_2}{m_1 - m_2}\right) m_1{}^t \\ & + \left(\frac{P_\Phi m_1 - P_{\Phi+1}}{m_1 - m_2}\right) m_2{}^t \end{aligned}$$

To solve the weekly model total demand at the beginning of the week is set equal to zero. This demand includes the interday equations (3) and the interweek equations (15) since both groups will trade at the opening price of the week. For the demand at the opening price of week u we have

$$(18) \qquad \begin{aligned} 0 = {} & -a\overset{*}{P}_{1-n+nu} + b + c(P_{n(u-1)} \\ & -P_{n(u-1)-1}) - aP'_u + b \\ & +c''(\bar{P}_{u-1} - \bar{P}_{u-2}) \end{aligned}$$

Since $\overset{*}{P}_{1-n+nu} = P'_u$ and $P_{n(u-1)} = \bar{P}_{u-1}$ equation (18) can be written as

$$(19) \qquad \begin{aligned} P_u' = {} & \frac{b}{a} + \frac{c}{2a}(\bar{P}_{u-1} - P_{n(u-1)-1}) \\ & + \frac{c''}{2a}(\bar{P}_{u-1} - \bar{P}_{u-2}) \end{aligned}$$

Definition A

$$\text{week } u \equiv \{\text{day } 1 - n + nu, \text{ day } 2 - n + nu, \cdots, \text{ day } nu\}$$

$$\text{week } u - 1 \equiv \{\text{day } 1 - n + n(u-1), \cdots, \text{ day } n(u-1)\}$$

$$\text{week } u - 2 \equiv \{\text{day } 1 - n + n(u-2), \cdots, \text{ day } n(u-2)\}$$

Using equation (9) we have

$$P_{1-n+nu} = \left(P'_u - \frac{b'}{a'}\right)e^{a'/c'} + \frac{b'}{a'} \tag{20}$$

Combining (19) and (20) we get

$$\begin{aligned} P_{1-n+nu} &= \left(\frac{\alpha+\beta}{2}\right)\overline{P}_{u-1} - \frac{\beta}{2}\overline{P}_{u-2} \\ &\quad - \left(\frac{\alpha}{2}\right)P_{n(u-1)-1} + H \end{aligned} \tag{21}$$

where

$$\alpha \equiv \frac{c}{a}e^{a'/c'}, \qquad \beta \equiv \frac{c''}{a}e^{a'/c'}$$

and H represents a collection of terms not depending upon u.

To obtain a difference equation with close of week price terms only, it is necessary to substitute for P_{1-n+nu} and $P_{n(u-1)-1}$ in equation (21). Using the daily closing difference equation P_{1-n+nu} can be replaced with $\overline{P}_u$ and $\overline{P}_{u-1}$. Setting $t=n$ and $\Phi=n(u-1)$ equation (17) becomes

$$\begin{aligned} P_{nu} &= \left(\frac{P_{1-n+nu} - P_{n(u-1)}\,m_2}{m_1 - m_2}\right)m_1^n \\ &\quad + \left(\frac{P_{n(u-1)}\,m_1 - P_{1-n+nu}}{m_1 - m_2}\right)m_2^n \end{aligned} \tag{22}$$

Setting $P_{nu}=\overline{P}_u$ and $P_{n(u-1)}=\overline{P}_{u-1}$, and substituting (21) into (22) we get equation (23).

To eliminate $P_{n(u-1)-1}$ from the equation it is necessary to replace it with $P_{1-n+n(u-1)}$ and $\overline{P}_{u-2}$ first, and then replace $P_{1-n+n(u-1)}$ with $\overline{P}_{u-1}$ and $\overline{P}_{u-2}$. This complication is encountered because equation (17) holds only when $\Phi+1$ and $\Phi+t$ are days in the same week. Setting $t=n-1$ and $\Phi=n(u-2)$ equation (17) becomes

$$\begin{aligned} P_{n(u-1)-1} &= \left(\frac{P_{1-n+n(u-1)} - m_2\overline{P}_{u-2}}{m_1 - m_2}\right)m_1^{n-1} \\ &\quad + \left(\frac{\overline{P}_{u-2}m_1 - P_{1-n+n(u-1)}}{m_1 - m_2}\right)m_2^{n-1} \end{aligned} \tag{24}$$

Substituting (24) into (23) and combining terms yields equation (25).

$$\begin{aligned} \overline{P}_u = &\left\{\left(\frac{m_1^n}{m_1-m_2}\right)\left(\frac{\alpha+\beta}{2} - m_2\right) + \left(\frac{m_2^n}{m_1-m_2}\right)\left(m_1 - \frac{\alpha+\beta}{2}\right)\right\}\overline{P}_{u-1} \\ &+ \left\{\left(\frac{m_1^n}{m_1-m_2}\right)\left(-\frac{\beta}{2}\right) + \left(\frac{m_2^n}{m_1-m_2}\right)\left(\frac{\beta}{2}\right)\right\}\overline{P}_{u-2} \\ &+ \left\{\left(\frac{m_1^n}{m_1-m_2}\right)\left(-\frac{\alpha}{2}\right) + \left(\frac{m_2^n}{m_1-m_2}\right)\left(\frac{\alpha}{2}\right)\right\}P_{n(u-1)-1} + H \end{aligned} \tag{23}$$

$$\begin{aligned} \overline{P}_u = &\left\{\left(\frac{m_1^n}{m_1-m_2}\right)\left(\frac{\alpha+\beta}{2} - m_2\right) + \left(\frac{m_2^n}{m_1-m_2}\right)\left(m_1 - \frac{\alpha+\beta}{2}\right)\right\}\overline{P}_{u-1} \\ &+ \left\{\left(\frac{m_1^n - m_2^n}{m_1-m_2}\right)\left(-\left(\frac{\beta}{2}\right) + \left(\frac{\alpha}{2}\right)\frac{m_1^{n-1}m_2}{m_1-m_2} - \left(\frac{\alpha}{2}\right)\frac{m_2^{n-1}m_1}{m_1-m_2}\right)\right\}\overline{P}_{u-2} \\ &+ \left\{\left(\frac{m_1^n - m_2^n}{(m_1-m_2)^2}\right)\left(-\frac{\alpha}{2}m_1^{n-1} + \left(\frac{\alpha}{2}\right)m_2^{n-1}\right)\right\}P_{1-n+n(u-1)} + H \end{aligned} \tag{25}$$

Using equation (17) once again, letting $t=n$ and $\Phi=n(u-2)$, and letting $P_{n(u-1)} = \overline{P}_{u-1}$ and $P_{n(u-2)}=\overline{P}_{u-2}$, and solving for $P_{1-n+n(u-1)}$ we obtain

$$
\begin{aligned}
P_{1-n+n(u-1)} &= \left(\frac{m_1 - m_2}{m_1^{n} - m_2^{n}}\right)\overline{P}_{u-1} \\
&\quad - \left(\frac{m_1 m_2^{n} - m_2 m_1^{n}}{m_1^{n} - m_2^{n}}\right)\overline{P}_{u-2}
\end{aligned} \tag{26}
$$

Substituting (26) into (25) we get equation (27). By rearranging and simplifying (27) we get equation (28) which is the weekly closing price difference equation when $(c/a)e^{a'/c'}$ does not equal four.

The necessary and sufficient conditions for stability of equation (28) are given by equation set (2) where a_1 and a_2 are defined as shown in equation (29).

When $(c/a)e^{a'/c'}$ equals four, the auxiliary equation (16) has real and equal roots. The general solution of equation (11) can be written as

$$
P_{\Phi+t} = \left(P_{\Phi} + (P_{\Phi+1} - 2P_{\Phi})\,\frac{t}{2}\right)2^{t} \tag{30}
$$

where 2 is the double root.

The solution of the weekly model when $(c/a)e^{a'/c'}$ equals four may be obtained by using equation (30) instead of equation (17), following the same procedure. The necessary and sufficient conditions for stability in this case are given by equation

---

$$
\begin{aligned}
\overline{P}_u = &\Bigg\{\left(\frac{m_1^{n}}{m_1 - m_2}\right)\left(\frac{\alpha+\beta}{2} - m_2\right) + \left(\frac{m_2^{n}}{m_1 - m_2}\right)\left(m_1 - \frac{\alpha+\beta}{2}\right) \\
&+ \left(-\left(\frac{\alpha}{2}\right)m_1^{n-1} + \left(\frac{\alpha}{2}\right)m_2^{n-1}\right)\left(\frac{1}{m_1 - m_2}\right)\Bigg\}\overline{P}_{u-1} \\
&+ \Bigg\{\left(\frac{m_1^{n} - m_2^{n}}{m_1 - m_2}\right)\left(-\frac{\beta}{2} + \left(\frac{\alpha}{2}\right)\frac{m_1^{n-1} m_2}{m_1 - m_2} - \left(\frac{\alpha}{2}\right)\frac{m_2 m_1^{n-1}}{m_1 - m_2}\right) \\
&+ \left(-\left(\frac{\alpha}{2}\right)m_1^{n-1} + \left(\frac{\alpha}{2}\right)m_2^{n-1}\right)\left(-\frac{m_1 m_2^{n} - m_2 m_1^{n}}{(m_1 - m_2)^2}\right)\Bigg\}\overline{P}_{u-2} + H
\end{aligned} \tag{27}
$$

$$
\begin{aligned}
\overline{P}_u = &\frac{1}{2(m_1-m_2)}\left\{(m_1^{n} - m_2^{n})(\alpha+\beta) + (m_1^{n-1} - m_2^{n-1})(-\alpha - 2m_1m_2)\right\}\overline{P}_{u-1} \\
&+ \frac{1}{2(m_1 - m_2)^2}\left\{(m_1^{n} - m_2^{n})\left[(-\beta)(m_1 - m_2) + (\alpha m_1m_2)(m_1^{n-2} - m_2^{n-2})\right]\right. \\
&\left. - \alpha m_1m_2(m_1^{n-1} - m_2^{n-1})^2\right\}\overline{P}_{u-2} + H
\end{aligned} \tag{28}
$$

$$
\begin{aligned}
a_1 &\equiv -\frac{1}{2(m_1 - m_2)}\left\{(m_1^{n} - m_2^{n})(\alpha+\beta) + (m_1^{n-1} - m_2^{n-1})(-\alpha - 2m_1m_2)\right\}; \\
a_2 &\equiv -\frac{1}{2(m_1 - m_2)^2}\left\{(m_1^{n} - m_2^{n})\left[(-\beta)(m_1 - m_2) + (\alpha m_1m_2)(m_1^{n-2} - m_2^{n-2})\right]\right. \\
&\qquad \left. -\alpha m_1m_2(m_1^{n-1} - m_2^{n-1})^2\right\}
\end{aligned} \tag{29}
$$

$$(31)\qquad \begin{aligned} a_1 &\equiv -\ 2^{n-1}\left\{\frac{n(\alpha+\beta)}{2} - \left(\frac{1}{4}\right)(n-1)(\alpha+8)\right\}\ ; \\ a_2 &\equiv -\ 2^{n-1}\left\{-\frac{\beta n}{2} + \alpha(n)(n-2)2^{n-2} - \alpha(n-1)^2 2^{n-2}\right\} \end{aligned}$$

---

set (2) where a_1 and a_2 are defined as shown in equation (31). Equation set (31) may also be derived by noting it is the limit of (29) when both m_1 and m_2 approach two.

The first result to be obtained concerns the effects of increasing the number of days in the week. It is necessary and sufficient for stability of the intraweek equation (11) that ρ be less than one, where

$$\rho = \max\left(|m_1|,\ |m_2|\right),$$

and m_1 and m_2 are the roots of the auxiliary equation (16). But the limits of a_1 and a_2 (as defined in (29)) when n becomes arbitrarily large are each zero if ρ is less than one. Therefore it may be concluded that the weekly model is stable regardless of the value of β if the intraweek system is stable and if the number of days in the week is sufficiently large. This result is plausible since if the number of days in the week becomes arbitrarily large, the model consists of one week of days and is equivalent to the daily model. Nevertheless, it is valuable to notice that if the intraweek system is stable, then increasing the length of the week by a sufficiently large amount will produce stability. More important, though, is what will be the effect upon stability of changing from an n day week to an $n+1$ day week. Alternatively, it may be of interest to examine what occurs when n is small. This will be examined briefly in what follows. Any result is obtainable, of course, for any specific n and any market by using (29) or (31) and the estimated values of α and β.

For the readers' convenience, the values of a_1 and a_2, and the corresponding stability conditions for the one-, two-, and three-day week will be written explicitly. Using (29) with n equal to one we obtain

$$a_1 = -\frac{\alpha+\beta}{2}$$

$$a_2 = \frac{\alpha+\beta}{2}$$

Using (2), necessary and sufficient conditions for stability when n equals one are

$$(32)\qquad -\frac{1}{2} < \frac{\alpha+\beta}{2} < 1$$

When n equals two

$$a_1 = -\frac{\alpha}{2}(\alpha+\beta) + \frac{3}{2}\alpha$$

$$a_2 = \frac{\alpha\beta}{2} + \frac{\alpha^2}{2}$$

since $m_1 = (\alpha/2) + (1/2)\sqrt{\alpha^2 - 4\alpha}$ and $m_2 = \alpha/2 - (1/2)\sqrt{\alpha^2 - 4\alpha}$ (from equation (16)). Necessary and sufficient conditions for stability of weekly closing prices when n equals two are

$$\alpha > -\frac{2}{3}$$

$$(33)\qquad \alpha^2 + \alpha\beta - \frac{3}{2}\alpha > -1$$

$$\alpha\beta + \alpha^2 < 2$$

When n equals three

$$a_1 = -\frac{\alpha^3}{2} - \frac{\alpha^2\beta}{2} + \frac{\alpha\beta}{2} + 2\alpha^2$$

$$a_2 = \frac{\alpha^2\beta}{2} - \frac{\alpha\beta}{2} + \frac{\alpha^3}{2}$$

Necessary and sufficient conditions for stability of the three-day week are

$$\begin{aligned} &\alpha^2\beta - \alpha\beta + \alpha^3 < 2 \\ &\alpha^3 + \alpha^2\beta - \alpha\beta - 2\alpha^2 > -1 \end{aligned} \tag{34}$$

The following theorems can be easily proved using conditions (32), (33), and (34).

THEOREM 1:

Given any α, there exists a β such that the one-day week is stable.

THEOREM 2:

The two-day week is unstable for $\alpha \geqq 2$ and $\alpha \leqq -2/3$.

THEOREM 3:

The three-day week is unstable for $\alpha \geqq \sqrt{3/2}$ and $\alpha \leqq -\sqrt{3/2}$.

These theorems indicate the possibility of stabilizing an unstable market when the intraweek system is unstable by shortening the length of the week (especially if interday expectations are extrapolative with c positive).

THEOREM 4:

If $\alpha > 1$ then two-day week stability implies one-day week stability.

PROOF:

$\alpha > 1$ implies $2-\alpha > (2/\alpha) - \alpha$. From (33) we have $\alpha\beta + \alpha^2 < 2$. Moreover, $\alpha\beta + \alpha^2 < 2$ implies $\beta < (2/\alpha) - \alpha$. Therefore $\beta < 2 - \alpha$ or $\alpha + \beta < 2$.

$\alpha > 1$ implies $(3/2) - (1/\alpha) - \alpha > -1 - \alpha$. From (33) we have $\alpha^2 + \alpha\beta - (3/2)\alpha > -1$ This implies $\beta > (3/2) - (1/\alpha) - \alpha$. Therefore $\beta > -1 - \alpha$ or $\alpha + \beta > -1$.

Since $-1/2 < (\alpha+\beta)/2 < 1$ are necessary and sufficient for one-day week stability as given in (32), the theorem is proved.

THEOREM 5:

If $\alpha > \sqrt{5} - 1$ and the intraweek closing price path has an oscillatory limiting behavior, then three-day week stability implies two-day week stability.

PROOF:

By (34) we know $\alpha^2\beta - \alpha\beta + \alpha^3 < 2$ and $\alpha^3 + \alpha^2\beta - \alpha\beta - 2\alpha^2 > -1$. If the intraweek closing price path has an oscillatory limiting behavior, then the roots of (16) are complex, or the root with largest absolute value is negative. If α is greater than or equal to four, the root with the greatest absolute value is positive. If $0 < \alpha < 4$ then the roots are complex. Since we assumed $\alpha > \sqrt{5} - 1$, we have $\sqrt{5} - 1 < \alpha < 4$. But $\sqrt{5} - 1 > 1$. Therefore $\alpha^2 - 5\alpha + 4 < 0$. This implies

$$\frac{-\alpha^3 + 2\alpha^2 - 1}{\alpha^2 - \alpha} > -\alpha - \frac{1}{\alpha} + \frac{3}{2}$$

But $\alpha^3 + \alpha^2\beta - \alpha\beta - 2\alpha^2 > -1$ implies

$$\beta > \frac{-\alpha^3 + 2\alpha^2 - 1}{\alpha^2 - \alpha}$$

Therefore $\beta > -\alpha - (1/\alpha) + (3/2)$ or $\alpha^2 + \alpha\beta - (3/2)\alpha > -1$. The condition $\alpha > \sqrt{5} - 1$ implies $\alpha > -2/3$. By (33), to complete the proof, we must prove $\alpha\beta + \alpha^2 < 2$. $\alpha > \sqrt{5} - 1$ implies $\alpha^2 + 2\alpha - 4 > 0$. But this in turn implies

$$\frac{2 - \alpha^3}{\alpha^2 - \alpha} < \frac{2}{\alpha} - \alpha$$

$\alpha^2\beta - \alpha\beta + \alpha^3 < 2$ implies

$$\beta < \frac{2 - \alpha^3}{\alpha^2 - \alpha}$$

Therefore $\beta < (2/\alpha) - \alpha$ or $\alpha\beta + \alpha^2 < 2$.

Theorems 4 and 5 indicate that if the intraweek system is moderately unstable, then eliminating a day from trading may not destabilize a stable market. Theorems 1 through 5 are examples of the types of results which would aid decision makers if the values of α and β cannot be accurately estimated.

VI. Conclusion

The major results of this paper include: Either intraday or interday behavior can compensate for the instability of the other; changing the length of the trading day can destabilize a stable market or stabilize an unstable market; changing the length of the trading week can also have similar stabilizing and destabilizing effects.

Several simplifications have been used in the preceding analysis which should be given explicit recognition. Perhaps most important is the avoidance of taking account of asset changes in the demand functions of the participants. Related to this is the absence of consideration of feasibility requirements (resource constraints) and of possible negativity of price. Another serious simplification is in the definition of stability. For the weekly model, in particular, it may be possible in special cases for both the intraweek system to be oscillatory divergent and the closing prices each week to converge to their equilibrium.

One final note of caution should be presented in connection with the results concerning the effect upon stability of varying the length of the day and week. It has been assumed, without explicit recognition, that all other parameters of the model would remain unchanged when these lengths are varied. This is only one of many possible approaches. Perhaps empirical work will indicate exactly how other variables may change with the length of the day x and of the week n. These relationships can then be applied in the analytical framework developed above.

Hopefully, the techniques developed in this paper can be applied to the many other markets which are characterized by scheduled discontinuous trading periods.

REFERENCES

K. J. Arrow and M. Nerlove, "A Note on Expectation and Stability," *Econometrica,* Apr. 1958, *26,* 297–305.

D. W. Bushaw and R. W. Clower, *Introduction to Mathematical Economics,* Homewood 1957.

A. C. Enthoven and K. J. Arrow, "A Theorem on Expectations and the Stability of Equilibrium," *Econometria,* July 1956, *24,* 288–93.

S. Goldberg, *Introduction to Difference Equations,* New York 1965.

J. Negishi, "The Stability of a Competitive Economy: A Survey Article," *Econometrica,* Oct. 1962, *30,* 635–69.

N. P. Obst, "Stability of a Competitive Market with Speculative Behavior," unpublished doctoral dissertation, Purdue Univ. 1970.

Officer Supply—The Impact of Pay, the Draft, and the Vietnam War

By STUART H. ALTMAN AND ROBERT J. BARRO*

In this study a model is developed to estimate a labor supply function of military officers. The model, discussed in Section I, assumes that in the absence of a draft the decision to enlist as an officer can be analyzed in the same context as a choice between two occupations; civilian and military. For a given level of military pay, the key determinants of the percentage who enlist are the distributions of alternative civilian earnings and of tastes for civilian relative to military life.

Section II shows that the introduction of a draft can be viewed as a mechanism which has the same effect as an increase in officer pay. Accordingly, any level of draft probability may be translated into an equivalent dollar increment in officer pay which would (in an all-volunteer system) have produced the same enlistment rate.

Section III carries out empirical estimation of the theoretical officer supply model using data on freshmen ROTC enrollment rates for 1960 to 1970. An important segment of the empirical analysis is the estimation of the changes in relative tastes for military life that resulted from the intensification of the Vietnam War.

In the final section the estimated coefficients are used to determine: 1) the extent to which officer supply is positively related to military earnings; 2) whether a premium has to be paid to the average civilian to volunteer for the officer corps, and if so, how much; 3) the effect of the Vietnam War on enlistments; and 4) the impact of changes in the probability of being drafted on potential officer supply.

I. The Decision to Enter Military Service Without a Draft

We shall investigate officer supply by utilizing a modified version of the occupational choice model originally suggested by Anthony Fisher (1969) in his study of military enlistments. According to this approach, each individual is viewed as having a choice between two "occupations," civilian or military officer. The relevant choice horizon is assumed to be limited to the short period corresponding to the length of initial duty obligation of an officer (two years for Army ROTC officers in the empirical section).[1] Because of the short time period involved, the choice of occupation is viewed here as a one-period decision problem.[2] In this single period context, the individual views each occupation as possessing pecuniary and nonpecuniary characteristics. The pecuniary characteristic, or monetary return, for the *i*th individual in the civilian occupation is denoted by: w_c^i. The monetary return in the officer occupation is denoted by: w_o. Given the rather uniform treatment of officers during the early years of service, this return is assumed to be the same for all individuals.

---

* Associate professor of economics at Brown University on leave as Deputy Assistant Secretary for Health Planning and Evaluation (HEW); and assistant professor of economics at Brown University, respectively. Both served as consultants to the President's Commission on an All-Volunteer Armed Force. A modified version of the model described in this paper was used by the Commission to estimate an officer supply function without a draft.

[1] Essentially, the choice between civilian and military is assumed to have no effect on opportunities subsequent to the initial period. This assumption is based on information obtained from Department of Defense surveys which show the similarity of the long-term earnings stream of veterans and non-veterans, standardized for age and education. See *The Report of the President's Commission on an All-Volunteer Armed Force*, ch. 12.

[2] If the entire stream of pecuniary and nonpecuniary earnings is considered for each occupation over a time period T, a sufficient set of conditions for collapsing the T-period problem to a one-period problem is: 1) the time pattern of civilian pecuniary earnings is the same for all individuals and corresponds to the pattern of the officer pecuniary earnings; (2) the time pattern of nonpecuniary characteristics in each occupation coincides with the pattern of the pecuniary characteristics; and (3) the *i*th individual's valuation of nonpecuniary char-

The nonpecuniary characteristic of the civilian occupation (measured in a positive physical unit) is N_c, which is valued (in money terms) by the *i*th individual as: $\rho_c^i N_c$, where ρ_c^i may be either positive or negative. In other words, the civilian occupation is viewed as possessing some physical characteristics, N_c, which are the same for all individuals, but individuals may vary in their valuations, ρ_c^i of these characteristics. Similarly, the officer occupation has nonpecuniary characteristic N_o, which is valued by the *i*th individual as $\rho_c^i N_o$.

Each individual evaluates the two occupations by calculating the sum of pecuniary and monetized nonpecuniary returns. Accordingly, the total monetized returns from the two occupations are:

$$\begin{aligned} M_c^i &= w_c^i + \rho_c^i N_c \\ M_o^i &= w_o + \rho_o^i N_o \end{aligned} \tag{1}$$

The individual is assumed to choose the occupation which provides the highest total return. Therefore, in this two-choice model, the decision to enlist as an officer is made if:

$$w_c^i + \rho_c^i N_c < w_o + \rho_o^i N_o \tag{2}$$

It is further assumed that, *ceteris paribus*, an individual's monetization factor for nonpecuniary returns is proportional to his monetary earnings in the respective occupation; that is:[3]

$$\rho_c^i = \tau_c^i w_c^i; \qquad \rho_o^i = \tau_o^i w_o \tag{3}$$

Defining the relative civilian/military taste factor:

$$(1 + \tau_i) = \frac{1 + \tau_c^i N_c}{1 + \tau_o^i N_o}, \tag{4}$$

and assuming that $(1+\tau_c^i N_c)$ and $(1+\tau_o^i N_o)$ are both positive for all *i*,[4] the officer enlishment condition of equation (2) may be simplified to:

$$w_c^i(1 + \tau_i) < w_o \qquad (1 + \tau_i > 0) \tag{5}$$

Therefore, $(1+\tau_i)$ in equation (5) determines the relative pay differential, w_o/w_c^i which is necessary to make the *i*th individual indifferent between the civilian and officer occupations. The differential exceeds one if the individual prefers the nonpecuniary characteristics of the civilian occupation, and is below one in the opposite case.

Among a sample of individuals, variations occur in w_c^i and $(1+\tau_i)$, but, we assume, not in w_o. Thinking of variations in w_c^i and $(1+\tau_i)$ as probability distributions the fraction of individuals in a group who enlist is given by:

$$\Phi = P_r[w_c^i(1 + \tau_i) < w_o] \tag{6}$$

where P_r may be interpreted as a cumulative probability.

For a given value of officer pay, the key

---

acteristics for each occupation is, *ceteris paribus*, proportional to the present value of the individual's pecuniary earnings in that occupation.

[3] Roughly speaking, the higher one's (money) income the greater the monetary value attached, *ceteris paribus* to a given physical quantity of nonpecuniary characteristic. At this point, equation (3) may be regarded as definitional. The substantive assumption, made below is that $(1+\tau_i)$ in equation (4) is distributed independently of w_c^i.

[4] If $(1+\tau_c^i N_c) \leq 0$, an individual with $(1+\tau_o^i N_o) > 0$ will enlist for any positive value of w_o. If $(1+\tau_o^i N_o) \leq 0$ an individual with $(1+\tau_c^i N_c) > 0$ will not enlist for any w_o if $w_c^i > 0$. These two extreme possibilities are excluded in this model.

elements which determine the enlistment rate in equation (6) are the distribution of civilian earnings and the distribution of tastes for civilian relative to military life. If w_c^i and $(1+\tau_i)$ can be treated as independently distributed, lognormal variates,[5] then $w_c^i(1+\tau_i)$ in equation (6) will also be distributed lognormally. The enlistment fraction of equation (6) may be expressed, in this case, as:[6]

$$\Phi = N\left[\frac{\mathit{log}\left(\dfrac{w_o}{\tilde{w}_c(1 \tilde{+} \tau)}\right)}{\sigma}\right] \tag{7}$$

where N is the cumulative standard normal density, $\tilde{w}_c$ is the median value of civilian earnings in the group under consideration, $(1 \tilde{+} \tau)$ is the median taste factor of the group, and σ is the standard deviation of $\mathit{log}\ [w_c^i(1+\tau_i)]$. For given values of $\tilde{w}_c$ and σ, the enlistment rate is indicated in equation (7) as a positive function of w_o and an inverse function of $(1 \tilde{+} \tau)$. Therefore, from the standpoint of officer procurement, enlistments in an all-volunteer environment may be increased either by raising w_o or by reducing $(1 \tilde{+} \tau)$, (i.e., by raising the nonpecuniary benefits of officer life).

The supply curve of equation (7) is depicted graphically in Figure 1. Considering Φ as a function of the adjusted pay ratio, $w_o/\tilde{w}_c(1 \tilde{+} \tau)$, the curve is positively sloped throughout. The enlistment rate increases with the pay ratio at an increasing rate for very low values of relative officer pay, that is, in the range[7]

$$0 \leq w_o/\tilde{w}_c(1 \tilde{+} \tau) < e^{-\sigma^2};$$

and at a decreasing rate for higher relative pay, that is, the range

$$w_o/\tilde{w}_c(1 \tilde{+} \tau) > e^{-\sigma^2}$$

Only the decreasing rate section of the function turns out to be relevant in the empirical study.

The elasticity of the enlistment rate with respect to the pay ratio is positive throughout and decreases with the pay ratio for all (positive) values of relative pay: $w_o/\tilde{w}_c(1 \tilde{+} \tau) > 0$. In the Klotz-Fisher interchange the elasticity is derived, and it is shown that the elasticity decreases throughout.[8]

---

[5] A rationale for the assumption of independence and the use of the lognormal distribution in this situation is given in Fisher (1969, pp. 242–43). While sufficient evidence exists to justify lognormality as an approximation to the distribution of civilian earnings, the argument is much more tenuous for $(1+\tau_i)$. Nevertheless, lacking additional information on the distribution of relative tastes, we followed Fisher by making such an assumption.

[6] As discussed by Fisher (1969, p. 243), the mean of log w_c^i is equal to $\mathit{log}\ (\tilde{w}_c)$, and the mean of $\mathit{log}\ (1+\tau_i)$ is equal to $\mathit{log}\ (1 \tilde{+} \tau)$, where a tilde indicates a median value. Since $\mathit{log}\ w_c^i$ and $\mathit{log}\ (1+\tau_i)$ are assumed to be independently, normally distributed, it follows that:

$$\frac{\mathit{log}\ [w_c^i(1+\tau_i)] - \mathit{log}\ (\tilde{w}_c(1 \tilde{+} \tau)]}{\sigma} \sim N(0,\ 1);$$

where

$$\sigma^2 = \mathrm{Var}\ (\mathit{log}\ w_c^i) + \mathrm{Var}\ (\mathit{log}\ (1+\tau_i))$$

Equation (7) follows from equation (6), using this standard normal distribution.

[7] The enlistment function is:

$$\Phi = N\left[\frac{\mathit{log}\left(\dfrac{w_o}{\tilde{w}_c(1 \tilde{+} \tau)}\right)}{\sigma}\right] = N\left[\frac{\mathit{log}\ (w)}{\sigma}\right],$$

where

$$w = \frac{w_o}{\tilde{w}_c(1 \tilde{+} \tau)}$$

The derivatives of Φ with respect to w are:

$$\frac{d\Phi}{dw} = \frac{1}{\sigma\sqrt{2\pi}w} e^{-1/2((\mathit{log}\ w)/\sigma)^2} > 0,$$

and

$$\frac{d^2\Phi}{dw^2} = \frac{-1}{\sigma\sqrt{2\pi}w^2}\left[\frac{\mathit{log}\ w}{\sigma^2} + 1\right] e^{-(1/2)((\mathit{log}\ w)/\sigma)^2}$$

The condition on the sign of the second derivative is:

$$\frac{d^2\Phi}{dw^2} \gtreqless 0 \quad \text{if } w \lesseqgtr e^{-\sigma^2}$$

[8] A different form of a supply function with a continuously decreasing elasticity was estimated by Altman (1969).

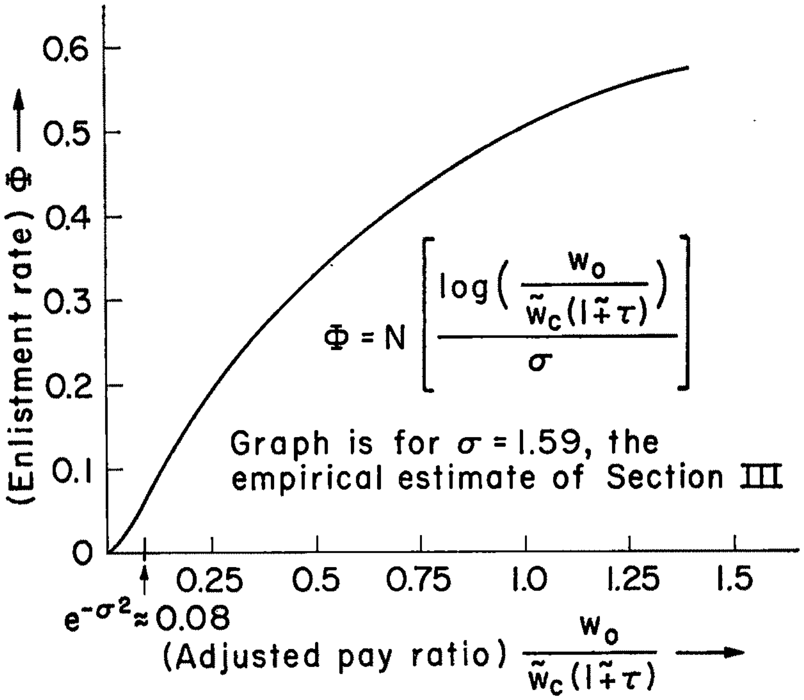


FIGURE 1. ENLISTMENT FUNCTION

II. The Effect of a Draft

With the introduction of a draft it is necessary to consider three occupations: civilian, officer, and draftee. As discussed below, there must be some distinction between being an officer and being a draftee if the introduction of a draft is to have any effect on the officer enlistment rate. In addition to the returns from the civilian and officer occupations, one must also calculate the return from the draftee occupation. For the following analysis we make the important simplifying assumption that the difference between the officer and draftee occupations can be captured solely as a pecuniary differential. Hence, the total monetized return from the draftee occupation is (see equations (1) and (3)):

$$(8) \qquad M_d^i = w_d + \rho_d^i N_o = w_d(1 + \tau_o^i N_o)$$

where w_d is draftee pay and is assumed to be equal for all individuals. The inclusion of the officer nonpecuniary term, $\tau_o^i N_o$, reflects the assumption that the officer and draftee occupations can be distinguished solely on pecuniary grounds.[9]

With the existence of a draft, an individual's choice is still between enlisting as an officer or remaining a civilian. The new element here is that there is some probability, p_i, of being drafted if one chooses to remain a civilian. Therefore, the civilian option amounts to the choice of a lottery where one obtains M_c^i with probability $(1-p_i)$, and M_d^i with probability p_i. Assuming that (von Neumann-Morgenstern) utility is linearly related to the monetized returns in each occupation —that is, that individuals are (at least approximately) risk-neutral in the choice among these occupations, the officer enlistment condition is:[10]

---

[9] In other words, the relative taste factor $(1+\tau_i)$ is specified in equation (4) as a nonpecuniary distinction between civilian and military life, but no taste distinction is made between different forms of military service (officer and draftee). A more complete model would add this additional taste factor, but the resulting complications, in terms of interpreting distribution functions, are considerable. Another useful extension would be the inclusion of a second civilian occupation which has a lower draft probability (occupational and teaching deferments), though possibly a lower value of M_c^i. See the discussion below.

$$(1 - p_i)M_c^i + p_iM_d^i < M_o^i$$

Using equations (1), (3), (4), and (8), the condition for volunteering as an officer may be expressed as (assuming $w_o > w_d$):

$$w_c^i(1 + r_i) < w_o + \frac{p_i}{1 - p_i}(w_o - w_d) \tag{9}$$

where $(1+r_i)$ is the civilian/military taste factor, defined in equation (4).

The enlistment condition in equation (9) indicates that the introduction of a draft ($p_i \neq 0$ for some i) has an impact on enlistment only if $w_o \neq w_d$. If officer and draftee pay are equivalent (in a generalized sense), all individuals who would opt for civilian life in the absence of a draft would still desire civilian life when a draft was introduced.[11] The essence of draft-induced enlistment is some distinction between the treatment of officers (enlistees) and draftees.

Again viewing w_c^i and $(1+r_i)$ as subject to probability distributions, the fraction of individuals who enlist in the presence of a draft is given by:

$$\Phi^d = P_r\left[w_c^i(1 + r_i) < w_o + \frac{p_i}{1 - p_i}(w_o - w_d)\right] \tag{10}$$

In equation (10) individuals are viewed as comparing taste-adjusted civilian pay, $w_c^i(1+r_i)$, with officer pay, w_o, augmented by a term to account for the effect of the draft: $[p_i/(1-p_i)](w_o-w_d)$. Among a group of individuals with a single value of $p_i=p$, and a fixed distribution of $w_c^i(1+r_i)$, the enlistment percentage increases with p if $w_o>w_d$, and increases with (w_o-w_d) if $p>0$. In other words, with the presence of a draft, the military has two vehicles in addition to increases in pecuniary and nonpecuniary pay which can increase the number of officer accessions. First, the probability of being drafted may be raised, and, second, the conditions of draftees, w_d, may be worsened.

The fraction of officer entrants in the presence of a draft (equation (10)) may be compared with those in the nondraft model (equation (6)). Since the left-hand variable in the probability statement, $w_c^i(1+r_i)$, is the same for both models, it is possible, without knowledge of the explicit distributions of w_c^i and $(1+r_i)$, to answer the following question. If there is currently a draft situation characterized by military wages w_o and w_d and draft probability p (which applies to all members of some group of individuals), and if Φ^d is the percentage of the eligible population which is voluntarily entering the military as officers, what increment in officer pay would be required to maintain the officer entrance rate if the draft were abolished? Since $\Phi = \Phi^d$ is required, the necessary increment in pay follows from equations (6) and (10):

$$\Delta w_o = \frac{p}{1 - p}(w_o - w_d)$$

As a numerical illustration, the data for 1964 are:[12] $p=.358$, $w_o - w_d = \$2444$/year. Therefore, the pay increment which just maintains the enlistment percentage when p is set equal to zero is: $\Delta w_o = \$1363$/year, which amounts to a 28 percent increase over 1964 first-term officer pay. Comparable figures for 1969 are: $p=.536$, $w_o - w_d = \$3619$, $\Delta w_o = \$4180$, or a 63 percent increment in pay. The greater percentage in the later year is accounted for by the higher percentage of officer enlistees who would otherwise be drafted.

---

[10] The possibility of volunteering for the draftee occupation is omitted, since $M_d^i < M_o^i (w_d < w_o)$ is assumed.

[11] This observation differs from results discussed by Fisher (1969, pp. 244, ff.).

[12] See Table 1 and the Appendix for these and additional data.

In order to determine the entire officer supply curve, it is necessary to specify the joint distribution of w_c^i, $(1+\tau_i)$ and p_i in equation (10). The analysis is facilitated if it is possible to segregate the total population into a few groups of equal draft probability. If a group with a single draft probability p is involved, and if w_c^i and $(1+\tau_i)$ are, as in Section I, independently lognormally distributed, then:

$$\Phi^d = N\left[\frac{\log \dfrac{\overset{*}{w_o}}{\tilde{w}_c(1 \tilde{+} \tau)}}{\sigma}\right] \tag{11}$$

where w_o^* is "effective" officer pay (explicit pay augmented by the effect of the draft), and is given from equation (10) by:

$$\overset{*}{w_o} = w_o + \frac{p}{1-p}(w_o - w_d) \tag{12}$$

III. Estimation of an Officer Supply Function

For most officer procurement programs in recent years the number of qualified applicants has exceeded the number of available positions. Hence, the number of officer volunteers is controlled by the needs of the military and does not reflect equality between supply and demand. One observable measure of officer supply is the enrollment rate of freshmen in voluntary Army ROTC I programs. Although enrollment in freshmen ROTC does not constitute entrance into military service, there is strong evidence that the factors which influence the officer volunteer rate have a similar influence on ROTC I enrollment.[13] It is necessary to use freshmen enrollment rates because it is only at this juncture and only for the Army program that all ROTC applicants are permitted to enter. The empirical measure of the enlistment rate is the ratio of four-year Army ROTC I enrollees (adjusted for the inclusion of advanced entrants to ROTC III beginning in 1966) to total male freshmen at sixty-five colleges.[14] The data refer to the fall of each fiscal year from 1960 to 1970. The schools included in the sample constitute an exhaustive list of those institutions which operated a voluntary Army ROTC program from 1956 to the present. Lack of important information for certain independent variables prevents the incorporation of observations prior to 1960. For each of the eleven sample years, the data are divided into separate averages for southern and nonsouthern schools, so that a total of twenty-two observations are available.

Because the officer supply function described by the normal distribution in equation (11) is non-linear, it is convenient to apply an indirect estimation procedure. First, the equation is approximated in a linear form, and the coefficients are estimated by standard regression analysis over a set of observations (which turned out to be the entire available sample) over which the linear approximation is satisfactory. Second, regarding the linear coefficients as elements of a Taylor expansion for the normal distribution around the mean of the sample observations, estimates of the distribution parameters are derived from the linear estimates. Finally, the estimated parameters of the non-linear form can be used to project officer supply in regions where the linear approximation is inappropriate.

---

[13] In an earlier study of officer supply, Alan Fechter also relied on freshmen ROTC data to estimate a cross-sectional labor supply function.

[14] Beginning in 1966 a select number of college juniors were permitted to enter a two-year Army ROTC program. An estimate of the impact of this program on freshmen enrollment in the four-year program was obtained by multiplying the number of two-year enrollees (two years hence) by a factor to account for normal attrition in the four-year program. This factor was calculated as a ratio of original freshmen enrollees to the number of students remaining in the four-year program at the start of the junior year. This adjustment was applied to freshmen enrollments in ROTC for the years 1966–70. The adjustments for 1969 and 1970 utilized the number of junior enrollees in the two-year program for 1970.

For the jth region at period t, the linear form is:

$$(13) \quad \Phi_{jt} = \alpha_1 + \alpha_2 \log \left(\frac{\overset{*}{w_o}}{\tilde{w}_c(1 \tilde{+} \tau)} \right)_{jt}$$

(j= South or non-South; t=1960, . . . , 1970) where Φ is the Army ROTC I enrollment rate, w_o^* is effective military pay, $\tilde{w}_c$ is median civilian pay, and $(1 \tilde{+} \tau)$ is the median taste factor for civilian relative to military life.

The use of equation (13) assumes for each (j, t) group-average observation that no intragroup variation occurs in w_o, w_d and p; which jointly determine w_o^* from equation (12). The assumption of uniform officer and draftee pay can be accepted as reasonable, but equality of draft probability among freshmen is clearly unreasonable. In general, this effect could be captured by describing some distribution of draft probabilities. This procedure has not been followed for two reasons. First, a continuous distribution of p's makes the analysis technically very difficult. Second, it appeared for the sample period considered that most of the variation in p could be described by measuring the relative proportion of freshmen contained in two polar groups: 1) an "a priori-deferred" group ($p=0$) consisting of most eventual physical deferees and some portion of eventual civilian deferees for whom, at the time of being a freshman, the draft probability was close to zero; and 2) an approximately homogeneous group of draft-prone individuals for whom the possibilities of escaping the draft involved mostly the likelihood of attaining a dependency deferment four years hence, or the chance of gaining entrance into a reserve unit.[15]

Operationally, the occupancy of the draft-prone group was estimated by first assuming that all physically deferred college graduates had sufficient knowledge of their condition as a freshman to consider their chances of being drafted as close to zero. From information obtained from a sample inventory of Selective Service registrants as of July 1964, 21 percent of male college graduates in each year were estimated to be physically deferred.[16] Among the total college graduate civilian deferred group (other than physical deferees) reaching age 26 in 1964, the distribution among categories was: student, 12.5 percent; occupational, 22.9 percent; dependency, 52.1 percent; other, 6.25 percent; I-A (not called), 6.25 percent. It was somewhat arbitrarily assumed that all of the student-occupational group and one-half of the remainder of eventual deferees were in the a priori deferred group. Hence, using the above distribution by categories, one-third of male college graduates who were eventually deferred on nonphysical grounds were classified in the draft-prone group.

The fraction of college graduates at the time they were freshmen considered to be in the draft-prone group equals:

$$(14) \quad \begin{aligned} \theta &= \frac{\text{Draft Prone}}{\text{Total Graduates}} \\ &= \frac{\frac{1}{3}X_c + X_o + X_d + X_r}{X_p + X_c + X_o + X_d + X_r} \end{aligned}$$

---

[15] The abolition of various categories of deferments and the adoption of a draft lottery system would alter the calculation of draft probability for the 1970's. These considerations do not affect estimation over the sample period of fiscal 1960 to fiscal 1970, but would affect estimation if data from the 1970's were eventually included.

[16] See *Review of the Administration and Operation of the Selective Service System.*

where,

X_p = physically deferred
X_c = civilian deferred (other than X_p)
X_o = entered as officers (less OCS entrants)[17]
X_d = draftees (plus enlistees and OCS officers)
X_r = reserve unit entrants

An annual tabulation of θ, shown in Appendix Table A-3, indicates a minimum value of 40 percent in 1965 and a maximum of 50 percent in 1969–70.

Those freshmen in the draft-prone group who do not enter ROTC (one component of X_o) have a probability p_d of being drafted, and a probability p_r of ending up in a reserve unit. The effective draft probability which gives full weight to p_d and half weight to p_r is:[18]

$$(15) \quad p = p_d + (\tfrac{1}{2})p_r = \frac{X_d + (\frac{1}{2})X_r}{(\frac{1}{3})X_c + X_d + X_r}$$

The relevant draft probability is that which is anticipated over the individual's years of draft eligibility (particularly at the time immediately following graduation from college). The empirical measure of p in this study gives half weight to the current (fiscal) year's draft percentage (which is partly anticipated at the time of enrollment in September of the freshmen year), and half weight to the previous year's figure.[19] Over the period 1960–70, p, tabulated annually in Table 1, has a minimum value of 0.31 in 1961–62, and a maximum value of 0.56 in 1970.

Since it was assumed that none of the a priori deferred group enrolls in Army ROTC, the revised enrollment equation is of the form:

$$(16) \quad \left(\frac{\Phi}{\theta}\right)_{jt} = \alpha_1 + \alpha_2 \, log \left[\frac{w_o^*}{\tilde{w}_c(1 \tilde{+} \tau)}\right]_{jt}$$

where θ is the draft-prone fraction given in equation (14), Φ is the percentage of ROTC enrollees out of total freshmen, and the value of p implicit in the w_o^* variable corresponds to the draft probability for members of the draft-prone group (equation (15)).

Except for the years 1960 to 1962, when a six-month ROTC program existed, completion of a four-year Army (nonscholarship) ROTC program implied a two-year active duty commitment (the same length as draftee service). Therefore, effective military pay is given from equation (12) by:[20]

---

[17] Officer candidate school entrants (OCS) and college graduate enlistees have been grouped with draftees, since most college graduates who enter service via these routes do so under imminent threat of the draft (or, after having been actually drafted).

[18] Entering the reserves involves 6 months of active service and a $5\frac{1}{2}$ year reserve obligation with periodic training stints, plus some possibility of being called to active duty. The inclusion of p_r with half weight in equation (15) assumes that participation in a reserve unit is viewed as equivalent to one year of draftee service.

[19] Adjustments to this measure have been made for 1962 and 1967. For fiscal 1962, it is clear that the major call-up later in the fiscal year was not anticipated in September 1961. Therefore, the value of p in fiscal 1961 was also used as a measure of p in 1962. For fiscal 1967 the overriding element in the draft probability calculation at the start of the freshman year in September 1966 was undoubtedly the already announced higher draft calls, so that p for 1967 is measured solely by the probability which actually prevailed in fiscal 1967.

[20] For the years 1960 to 1962, an ROTC enrollee's commitment was 6 months of active duty (with the possibility, actually exercised in 1962, of being retained for a longer period) plus $5\frac{1}{2}$ years of reserve obligation. We have assumed that the 6 month program was actually viewed as equivalent to a one-year active duty commitment as an officer, from which it follows that the enlistment criterion for this period is:

$$\left(\frac{1}{2}\right) w_o + \left(\frac{1}{2}\right) w_c^i(1+\tau_i) > (1-p_i) w_c^i(1+\tau_i) + p_i w_d$$

From the above criterion, effective military pay for the 6 month program years may be calculated as:

$$w_o^* = w_o + \left[\frac{p}{0.5 - p}\right](w_o + w_d)$$

Except for this modification of effective officer pay, the 6 month program years (1960–62) are treated in the same manner as the other years of the sample.

$$\overset{*}{w_o} = w_o + \frac{p}{1-p}(w_o - w_d)$$

where p is calculated from equation (15), and the empirical pay measures (contained in the Appendix) are:

w_o = average annual pay for first-term Army officers (over the first two years of service);

w_d = average annual pay for first-term draftees (over the two-year first-term period).

The measure of $\bar{w}_c$ in equation (16) is the average yearly earnings of college graduates over the two-year period subsequent to graduation.[21] The division of $\bar{w}_c$ according to South/non-South average pay is based on the single year, 1961, for which $\bar{w}_{c,s} = .910 \cdot \bar{w}_c$, and $\bar{w}_{c,n} = 1.038 \cdot \bar{w}_c$ (where $\bar{w}_c$ is national average pay).

An important element in equation (16) is the median civilian/military taste factor: $\log(1\bar{+}\tau)_{jt}$. A major time-related influence on civilian/military tastes over the period 1960–70 is the Vietnam War. Particularly with intensification of the conflict in recent years, there is strong presumptive evidence of increased dissatisfaction with the military, as well as a direct increase in a military man's chance of becoming a casualty. Presumably, higher dissatisfaction and increased casualty probability should result in a positive relationship between taste for civilian life $(1\bar{+}\tau)$ and the intensity of the war. The measure of war intensity employed in this study is based on the average casualty rate for junior Army officers over the preceding fiscal year (K_{t-1}).

In order to allow for a gradual response of tastes to war intensity, we have employed the partial adjustment relation:

$$K_t^* = \gamma K_{t-1} + (1-\gamma)K_{t-1}^* \qquad (0 \le \gamma \le 1) \tag{17}$$

where K_t^* can be equivalently written as a distributed lag in $K_{t-1}, K_{t-2}, K_{t-3}, \ldots,$ with exponentially declining weight: γ, $\gamma(1-\gamma)$, $\gamma(1-\gamma)^2, \ldots.$ The long-run casualty measure, K^*, has been used to explain tastes in the form:

$$\log(1\bar{+}\tau)_{jt} = \beta_1 + \beta_{2,j}K_t^*(\gamma) + \beta_3 I_s \qquad (j = s,.n) \tag{18}$$

Equation (18) allows for separate casualty-rate coefficients for South and non-South, since there is some a priori indication that the average individual in the South is more favorable (i.e., less opposed) to the war than the average individual in the non-South. In this case we expect: $\beta_{2,n} > \beta_{2,s} > 0$. The equation also includes the dummy variable I_s which equals one for South and zero for non-South. If (independently of the war) individuals in the South have a relative preference for military life over those in the non-South, then $\beta_3 < 0$.

Unfortunately, because K_t is essentially a time trend over the period of observation (Appendix Table A-2), it is impossible to estimate the distributed-lag coefficient of K^* from the sample data with any precision. Accordingly, the final estimate, $\hat{\gamma} = 0.25$/year, (and, therefore, the implied K^* series) is based on prior ideas of a plausible order of magnitude, rather than on empirical estimation. While variations in $\hat{\gamma}$ have a negligible impact on estimation over 1960–70,[22] shifts in this coef-

---

[21] The source of this data is the College Placement Council, June 1960–June 1969.

[22] The regression results for various values of $\hat{\gamma}$ are (see equations (19) and (20), below):

$\hat{\gamma}$	$\hat{a}_2$ (Pay (Coefficient)	Standard Error of Estimate
1.0	.226 (.036)	.031
0.5	.231 (.035)	.030
0.25	.233 (.034)	.029
0.10	.235 (.034)	.029

The numbers in parentheses are standard errors for $\hat{a}_2$, conditional on the set value of $\hat{\gamma}$. The table indicates that the estimated pay coefficient and regression fit are extremely insensitive to variations in $\hat{\gamma}$ over the range of 0.10 to 1.0.

ficient have a substantial impact on enlistment rate projections in the 1970's during which K_t hopefully reverses its upward trend. On the other hand, the estimated coefficient of the pay variable is very insensitive to the choice of $\hat{\gamma}$.

Combining equations (16) and (18), the form which was used for estimation is:

$$\left(\frac{\hat{\Phi}}{\theta}\right)_{jt} = \hat{a}_1 + \hat{a}_2\, log\, (w_o^*/\tilde{w}_c)_{jt} + \hat{a}_{3j}K_t^*(\hat{\gamma}) + \hat{a}_4 I_s \tag{19}$$

where the a coefficients are related to the coefficients of equations (16) and (18) by:

$$a_1 = \alpha_1 - \alpha_2\beta_1$$
$$a_2 = \alpha_2$$
$$a_{3,j} = -\alpha_2\beta_{2,j}$$
$$a_4 = -\alpha_2\beta_3$$

The expected signs of the coefficients are: $a_1>0$; $a_2>0$; $a_{3,n}<a_{3,s}<0$; $a_4>0$.

As indicated at the beginning of this section, the data used to estimate equation (19) comprise twenty-two total observations—eleven annual observations from 1960 to 1970 divided into South and non-South regions. In the initial estimation in the form of equation (19), $\hat{a}_4$ was insignificantly different from zero. Since there was no strong a priori belief in the importance of the South/non-South dummy variable, it was excluded from the final estimating equation.[23] The estimated regression coefficients with standard errors in parentheses (using natural logarithms) are:

$$\left(\frac{\hat{\Phi}}{\theta}\right) = \underset{(.009)}{.378} + \underset{(.034)}{.233}\, log\, (w_o^*/\tilde{w}_c) - \underset{(.050)}{.448}\, K^*(.25) \text{ South} - \underset{(.048)}{.554}\, K^*(.25) \text{ non-South} \tag{20}$$

$R^2 = .915$, standard-error-of-estimate $= .029$,

$D.W. = 2.57$, Mean$\left(\frac{\Phi}{\theta}\right) = .351$

The coefficient of the relative pay variable, $log\,(w_o{}^*/\tilde{w}_c)$ is, as expected, significantly positive. The casualty coefficients are both significantly negative, with the non-South coefficient (significantly) more negative than the South.[24] In other words, there is some evidence that the college population of the non-South has responded more negatively to the war than those in the South, though a substantial negative response is indicated in both regions.

The standard-error-of-estimate is about 8 percent of the mean of the dependent variable, and the Durbin-Watson statistic indicates absence of (positive) serial correlation in the residuals.[25]

The theory suggests that only the ratio of effective military to civilian pay is relevant for enlistments, and the above estimation has been carried out subject to this constraint. A test of this aspect of the

---

[23] The regression which includes the South/non-South dummy is:

$$\left(\frac{\hat{\Phi}}{\theta}\right) = \underset{(.011)}{.375} + \underset{(.037)}{.229}\cdot log\, (w_o^*/\tilde{w}_c) - \underset{(.055)}{.455}\cdot K^*(.25)_{\text{South}} - \underset{(.055)}{.545}\cdot K^*(.25)_{\text{non-South}} + \underset{(.017)}{.006}\cdot I_s$$

$R^2 = .916$, standard-error-of-estimate $= .030$, $D.W. = 2.52$.

While the sign of the I_s coefficient is positive, it does not differ significantly from zero. Therefore, the sample information does not substantiate a constant civilian/military taste differential for South versus non-South. In other words, for equal values of w_o^* and $\tilde{w}_c$, and with $K^*=0$, the enlistment rates in the South should equal those in the non-South.

[24] A regression which constrains the South/non-South casualty coefficients to be equal (which implies an equal reaction of tastes in each region to the war) yields:

$$\left(\frac{\hat{\Phi}}{\theta}\right) = \underset{(.009)}{.376} + \underset{(.035)}{.246}\cdot log\, (w_o^*/\tilde{w}_c) - \underset{(.040)}{.504}\cdot K^*(.25)$$

$R^2 = .901$, standard-error-of-estimate $= .031$, $D.W. = 2.24$

A test for equality of the K^*—South and K^*—non-South coefficients yields the t-statistic: $t_{18} = 1.78$. Therefore, a one-sided test of the null hypothesis, non-South = South, is rejected at the 95 percent level in favor of non-South < South < 0).

[25] The Durbin-Watson statistic is calculated on the basis of two separate groupings of eleven time-based observations each.

model is afforded by running the unconstrained regression:

$$\left(\frac{\hat{\Phi}}{\theta}\right) = \hat{b}_1 + \hat{b}_2\, log\,(w_o^*) + \hat{b}_3\, log\,(\tilde{w}_c) + \hat{b}_4\, K^*(.25) \text{ South} + \hat{b}_5 K^*(.25) \text{ non-South}$$

and testing $b_2 = -b_3$. The estimated equation is:

$$(21) \quad \left(\frac{\hat{\Phi}}{\theta}\right) = \underset{(.948)}{.036} + \underset{(.048)}{.245} \cdot log\,(w_o^*) - \underset{(.083)}{.206} \cdot log\,(\tilde{w}_c) - \underset{(.099)}{.478} \cdot K^*(.25) \text{ South} - \underset{(.123)}{.595} \cdot K^*(.25) \text{ non-South}$$

$R^2 = .916$, standard-error-of-estimate $= .030$, $D.W. = 2.63$

The null hypothesis, $b_2 = -b_3$, corresponds to a statistic: $t_{17} = 0.37$. Therefore, the hypothesis that only the ratio of pay is relevant for the enrollment rate is accepted. This result provides substantial statistical support for the theoretical model.

Parameters of the Normal Distribution

The regression fit of equation (20) is intended as a linear approximation to the normal curve:

$$(22) \quad \left(\frac{\Phi}{\theta}\right)_{jt} = N\left[\frac{log\,(\overset{*}{w}_o/\tilde{w}_c) - log\,(1 \tilde{+} \tau)}{\sigma}\right]_{jt}$$

where $log\ (1 \tilde{+} \tau)_{jt} = \beta_1 + \beta_{2,j} K_t^*$; and σ, the square root of the income plus taste variance, is assumed constant across j and t. Viewing equation (19) as the first two terms of a Taylor series expansion about the sample mean: $(\overline{\Phi/\theta}) = .351 = N(-.382)$, we have:

$$\left(\frac{\Phi}{\theta}\right)_{jt} \approx \left(\frac{\overline{\Phi}}{\theta}\right) + n(-.382) \cdot \left[\frac{log\,(w_o^*/\tilde{w}_c) - log\,(1 \tilde{+} \tau)}{\sigma} + .382\right]_{jt}$$

where n is the standard-normal density. Therefore, the linear coefficients of equation (19) correspond to the normal parameters as follows:

$$a_1 = \left(\frac{\overline{\Phi}}{\theta}\right) + n(-.382)\,[.382 - \beta_1/\sigma]$$

$$a_2 = \frac{n(-.382)}{\sigma}$$

$$a_{3,s} = \frac{-n(-.382)}{\sigma} \cdot \beta_{2,s}$$

$$a_{3,n} = \frac{-n(-.382)}{\sigma} \cdot \beta_{2,n}$$

Using the estimates of the a's from equation (20) we obtain the point estimates:

$$\hat{\sigma} = 1.59$$

$$(23) \quad log\,(1 \tilde{+} \tau)_{s,t} = .494 + 1.92\, \overset{*}{K}_t(.25)$$

$$log\,(1 \tilde{+} \tau)_{n,t} = .494 + 2.38\, \overset{*}{K}_t(.25)$$

At $K^* = 0$, the estimated taste factor is: $(1 \tilde{+} \tau) = 1.64$ (for both South and non-South). That is, at zero casualty rates, civilian pay is, on the average among the draft-prone group, multiplied by 1.64, and is then compared with (effective) military pay. At the positive values of K^* the taste factor is above 1.64 as indicated by equation (23), with the non-South taste factor responding more strongly to increased casualties than the south. For example, at the peak K^* value in 1970, the taste factors are $(1 \tilde{+} \tau)_s = 4.57$ and $(1 \tilde{+} \tau)_n = 5.84$.

Using equations (22) and (23), we may write:

$$\hat{\Phi}_{jt} = \theta_t \cdot N$$

$$(24) \qquad \cdot \left[\frac{\log (w_o^*/\bar{w}_c) - \log (1 + \tau)}{1.59}\right]_{jt}$$

where $\log\ (1 \tilde{+} \tau)_{jt}$ is given in equation (23). An estimate of the overall enlistment rate is given by:

$$(25) \qquad \hat{\Phi}_t = v_{s,t}\hat{\Phi}_{s,t} + v_{n,t}\hat{\Phi}_{n,t}$$

where v_s and v_n are the relative fractions of college freshmen in the South and non-South, respectively (Appendix Table A-2).

Using the parameter values of equation (23), estimated sample values of Φ have been calculated for 1960 to 1970, and are indicated along with actual values of Φ in Table 1. The table also includes sample period estimates of $\Phi(p=0)$—the enlistment percentage obtained from equation (25) with p set to zero in the expression for w_o^* in equation (12). Given estimates of Φ and $\Phi(p=0)$, the estimated draft induced fraction, $[\hat{\Phi}-\hat{\Phi}(p=0)]/\hat{\Phi}$, is calculated for each year.[26]

The estimated equation is also used to project enrollment rates in 1971 and later years, first on the assumption that the draft is maintained at the 1964 level ($p = .358$), and, second, on the assumption that the draft is abolished ($p=0$). These values are also contained in Table 1. It is assumed throughout that the ratio of military to civilian pay remains at the 1970 level ($w_o/\bar{w}_c = .741$, $w_d/\bar{w}_c = .382$).

TABLE 1—ARMY ROTC I ESTIMATED ENROLLMENT RATES

Fiscal Year	p	Φ	$\hat{\Phi}(p)$	$\hat{\Phi}(p=0)$	Draft Induced Fraction·	$\hat{\eta}(p=0)$
1960	.335	.193	.213	.139	.35	.71
1961	.308	.198	.195	.131	.33	.72
1962	.308	.196	.199	.136	.32	.73
1963	.387	.166	.150	.125	.17	.75
1964	.358	.147	.154	.128	.17	.75
1965	.328	.123	.132	.111	.16	.76
1966	.349	.151	.158	.131	.17	.76
1967	.525	.189	.166	.119	.28	.80
1968	.520	.138	.145	.101	.30	.85
1969	.536	.121	.123	.080	.35	.96
1970	.559	.075	.091	.054	.41	1.09
1971	.358	—	.064	.047	.27	1.09
1972	.358	—	.071	.054	.24	1.05
1973	.358	—	.087	.068	.22	.97
1974	.358	—	.100	.079	.21	.92
1975	.358	—	.111	.088	.21	.89
1980	.358	—	.137	.111	.19	.80
∞	.358	—	.146	.120	.18	.78

Notes: p is the draft probability of the draft-prone group, the same value has been used in each year for South and non-South regions; Φ is the actual fraction of freshmen enrolled in Army ROTC I; $\hat{\Phi}(p)$ is the estimated value of Φ (equation (25)); $\hat{\Phi}(p=0)$ is the estimated enrollment rate at $p=0$; draft-induced fraction $=[\hat{\Phi}(p)-\hat{\Phi}(p=0)]/\hat{\Phi}(p)$; $\hat{\eta}$ is the elasticity of the enrollment rate with respect to the military/civilian pay ratio (equation (26)).

Pay Elasticities

The impact of shifts in (military or civilian) pay on enlistments is best described by pay elasticities. From the normal distribution of equation (22), the elasticity of the enlistment rate with respect to the military/civilian pay ratio is:

$$\eta = \frac{\partial \log \Phi}{\partial \log (w_o^*/\bar{w}_c)}$$

$$(26) \qquad = \frac{1}{\sigma} \cdot \frac{n\left[\dfrac{\log (w_o^*/\bar{w}_c) - \log (1 \tilde{+} \tau)}{\sigma}\right]}{N\left[\dfrac{\log (w_o^*/\bar{w}_c) - \log (1 \tilde{+} \tau)}{\sigma}\right]}$$

where n is the standard normal density and N is the cumulative normal.[27] Estimated values of σ and $log\ (1+\tau)$ are those indicated in equation (23). Pay elasticities corresponding to estimated and projected enrollment rates (without a draft) have been calculated from equation (26) and are entered in Table 1.

---

[26] An alternative method for calculating the draft-induced fraction of officer volunteers is provided by direct survey results from the Department of Defense Survey of Active Duty Personnel (see *Report of the President's Commission on an All-Volunteer Armed Force*, 1970). This survey information was utilized in conjunction with the estimated values from this paper in projecting all-volunteer officer accession rates for the Commission Report (see Altman and Barro).

IV. Implications of Results

The model described above has been used as a part of the staff study for the President's Commission on an All-Volunteer Armed Force to estimate officer accessions in the 1970's under a draft-free system. Whether the projected ROTC enrollment rates for the 1970's indicated in Table 1 are sufficient to satisfy the Army's "requirements" without a draft depends on many factors other than those discussed in this paper—e.g., expected size of the Armed Forces, attrition in the ROTC program, accessions from academies and other officer recruitment programs, and career officer retention rates. Rather than speculate here on analysis which will be included in the staff's *Appendix Report* (Altman and Barro), we would rather highlight those conclusions that follow directly from the analysis of ROTC enrollments.

The occupational-choice model of the individual developed in this study includes a taste factor $(1+\tau_i)$ for civilian relative to military life. For the years prior to the intensification of the Vietnam War (1960–66), the estimated model suggests that, on the average, college freshmen had a negative taste for military life and required a premium to enter military service. Even at zero casualty rates, equation (23) indicates that civilian earnings should be multiplied by a factor of about 1.6 before comparing it with military pay. This conclusion does not necessarily imply that a premium above civilian pay is required to recruit an all-volunteer officer corps. The necessary level of officer pay depends on the fraction of the relevant population (which is itself about half of the total male freshmen population) which is needed to meet military demands. As the required fraction approaches half of the relevant population, the necessary pay premium approaches 64 percent of average civilian earnings.

A further implication of the analysis is that the intensification of the Vietnam War led to a sharp increase in taste for civilian over military life. For students in northern schools, the multiplicative taste factor reached 5.8 during the peak war year of 1970. Students in southern schools were somewhat less negatively affected by the war, with the 1970 taste factor reaching 4.6 in this region.

As shown in Table 1, the estimated ROTC I enrollment rate, $\hat{\Phi}$, declined from .132 in 1965 to .091 in 1970, or a drop of 31 percent. During this same period the estimated probability of a college graduate in the relevant population being drafted increased from .33 to .56. Had the draft probability remained at the 1965 level, the decline in $\hat{\Phi}$ would have been 48 percent instead of 31 percent. Column 4 of Table 1 contains projections of Φ beyond 1970 on the assumption that draft and casualty conditions return to prewar levels, and that relative military to civilian pay is maintained at the 1970 level. Although the intensity of the war is assumed to diminish rapidly, the distributed-lag effect of casualties on tastes and the drop in draft probability produce a significant decline

---

[27] A linear approximation to η from equation (20) is:

$$\eta \approx \frac{.233}{(\Phi/\theta)}$$

However, this approximation departs significantly from the estimates of equation (26) if one moves some distance from the sample mean.

in $\hat{\Phi}$ in 1971. For the period subsequent to 1971, there is a gradual increase in enrollments because of the gradual attenuation of the war's impact. By 1975, $\hat{\Phi}(p)$ is 73 percent above the estimated 1971 rate, but still 16 percent below the prewar 1965 rate. Fortunately for officer recruitment, the number of freshmen in 1975 in colleges included in our sample is estimated to be 57 percent above the 1965 level (see Altman and Barro).

The magnitude of the impact of the draft on the voluntary recruitment rate is an important question which must be addressed by any statistical study of military manpower. In this study the draft has been incorporated into the analysis as a mechanism for increasing effective officer pay, i.e., as a pay increment: $\Delta w_o = [p/(1-p)](w_o - w_d)$. Table 1, column 5, contains estimates for 1960–70 on ROTC enrollments which would have occurred in an all-volunteer system ($p=0$). The estimated "draft induced" fraction of volunteers is shown in the following column. In the years 1960–62, the estimated fraction of draft-induced volunteers is quite high—over 30 percent. In this period an individual could enter military service as an officer for only six months, as opposed to the two-year commitment in later years. With the end of the six-month program, the enrollment rate and the draft-induced fraction declined. By 1965, the fraction of draft-induced volunteers is estimated to be only 16 percent. Thereafter, with rising draft probability and a declining taste for the military (reduced relative number of true voltunteers), the fraction of draft-induced volunteers rises to a peak of 41 percent in 1970. In the projection period beyond 1970 (if the draft is maintained at the 1964 level), the decline in draft probability and war intensity should eventually reduce the draft-induced fraction to below 20 percent.

The pay elasticity of enrollment under the all-volunteer system is indicated in column 7 of the table. The elasticity has a peak value of 1.1 during the trough enrollment years of 1970–71, and declines steadily during the projection period, 1970–80, as enrollments are rising. The elasticity in 1980 is 0.8.[28] The inverse relation between pay elasticity and the enrollment rate reflects the general property of the supply function derived earlier.

In summary, this study suggests that an aggregative occupational-choice model can be used to explain the proportion of male college freshmen who volunteer to be military officers. The model can be adjusted to incorporate the existence of a draft by raising the effective level of military pay. When the parameters of the model are estimated using Army ROTC data for the period 1960–70, they indicate that: 1) the officer volunteer rate is positively related to the ratio of military to civilian earnings; 2) with the low percentage who would enlist without a draft in 1970–71, the elasticity is slightly higher than 1.0, but declines as the volunteer rate increases; 3) even in peacetime a premium has to be paid above civilian earnings to recruit half the relevant population; and 4) this premium increased quite sharply because of the Vietnam War.

---

[28] These estimates are considerably smaller than the constant elasticity of 1.98 reported by Fechter and by Altman and Fechter in previous studies of Army officer supply. They are, however, in line with the range of elasticity estimates calculated for enlisted personnel by Altman and by Fisher (1969). Benjamin Klotz, pp. 975–77, calculates elasticities on the basis of a priori reasoning and concludes that an elasticity greater than one " . . . is theoretically indicated for most plausible situations." His theoretical conclusion conflicts with our estimated values (which lead to elasticity estimates below 1.0 in some plausible situations) because our estimated standard error of $w_c^i(1+\tau_i)$—$\hat{\sigma}=1.59$—is above the values which Klotz considered for his "guesstimates."

APPENDIX

TABLE A-1—VALUES OF MILITARY AND CIVILIAN PAY

Fiscal Year	$\tilde{w}_c$	w_o	w_d	w_o^*	$w_o^* - w_o$
1960	5916	4547	2336	9036	4489
1961	6186	4547	2336	8094	3547
1962	6456	4547	2336	8094	3547
1963	6750	4547	2336	5943	1396
1964	7008	4805	2361	6168	1363
1965	7218	5004	2342	6303	1299
1966	7440	5659	2472	7367	1708
1967	7956	5972	2751	9532	3560
1968	8526	6538	2901	10478	3940
1969	9090	6623	3004	10803	4180
1970	9783	7247	3212	12361	5114

Notes: $\tilde{w}_c$ is average annual civilian pay of college graduates for the two-year period after graduation; w_o is annual average first-term (two years) Army officer pay; w_d is annual average first-term (two years) enlisted men's pay; w_o^* is effective officer pay (equation (12)); $w_o^* - w_o$ measures the necessary increment in officer pay which would just offset the impact of going off the draft in each year.

TABLE A-2—REGRESSION VARIABLES

Fiscal Year	K_{t-1}	$K_t^*(.25)$	$w_o^*/\tilde{w}_c$	$w_o/\tilde{w}_c$	$w_d/\tilde{w}_c$	v_s	v_n
1960	≈ 0	≈ 0	1.53	0.77	0.39	(.269)	(.731)
1961	≈ 0	≈ 0	1.31	0.74	0.38	(.269)	(.731)
1962	.010	.0025	1.25	0.70	0.36	.269	.731
1963	.010	.0044	0.88	0.67	0.35	.272	.728
1964	.030	.0108	0.87	0.69	0.34	.269	.731
1965	.093	.0314	0.87	0.69	0.32	.259	.741
1966	.180	.0686	0.99	0.76	0.33	.301	.699
1967	.328	.133	1.20	0.75	0.35	.297	.703
1968	.444	.211	1.23	0.77	0.34	.285	.715
1969	.772	.351	1.19	0.73	0.33	.313	.687
1970	1.081	.534	1.26	0.74	0.33	.301	.699
1971	(.54)	(.536)	(0.97)	(0.74)	(0.33)	(.301)	(.699)
1972	(.315)	(.481)	(0.97)	(0.74)	(0.33)	(.301)	(.699)
1973	(.09)	(.383)	(0.97)	(0.74)	(0.33)	(.301)	(.699)
1974	(.09)	(.310)	(0.97)	(0.74)	(0.33)	(.301)	(.699)
1975	(.09)	(.255)	(0.97)	(0.74)	(0.33)	(.301)	(.699)
1980	(.09)	(.129)	(0.97)	(0.74)	(0.33)	(.301)	(.699)
∞	(.09)	(.09)	(0.97)	(0.74)	(0.33)	(.301)	(.699)

Notes: K_{t-1} is the percentage casualty rate of junior Army officers over the previous year; K^* (.25) is a distributed lag of K_{t-1} with adjustment coefficient of .25; v_s is the fraction of southern and v_n that of nonsouthern freshmen in the overall observed college population.

Parentheses indicate estimated rather than actually observed values.

Table A-3—Estimation of the Draft Prone Proportion of Male College Graduates

(Numbers in Thousands)

Fiscal Year	X_t	X_p	X_o	X_d	X_r	X_c	θ
1960	222.5	45.8	24.7	14.4	19.8	117.8	.441
1961	223.1	46.1	23.6	10.1	20.5	122.8	.426
1962	223.4	46.1	25.7	25.5	14.1	112.0	.459
1963	228.4	47.2	23.3	18.3	16.9	122.7	.436
1964	239.1	49.4	20.6	17.7	27.7	123.7	.449
1965	263.1	54.3	19.6	16.0	18.0	155.1	.401
1966	279.8	57.8	26.4	26.2	31.9	137.5	.466
1967	301.1	62.2	27.0	53.6	14.3	144.0	.475
1968	324.2	66.9	26.9	57.6	9.8	162.9	.459
1969	359.7	74.3	32.8	69.6	26.7	156.3	.503
1970							(.503)

Notes: The numbers in the table refer to June graduates, immediately before the indicated fiscal year.

X_t = total male college graduates. Includes all male baccalaureate degrees taking less than five years. Excludes other first professional degrees (*Projections of Educational Statistics*).

X_p = individuals physically deferred for military service. Assumes a physical deferment rate for college graduates of .2065 (Selective Service Inventory, July, 1964).

X_o = newly commissioned college graduate military officers—accessions from academies, ROTC, and other college programs (except OCS) including M.D.'s—derived from unpublished Defense Department statistics.

X_d = college graduate draftees (including enlistees and OCS officers), derived from unpublished Defense Department statistics.

X_r = college graduate reserve enlistees, derived from unpublished Defense Department statistics.

X_c = nonphysically deferred nondrafted college graduates who did not enlist as an officer or enter the reserves: $X_c = X_t - (X_p + X_o + X_d + X_r)$.

$\theta = (1/3X_c + X_o + X_d + X_r)/X_t$, equation (14). The same value of θ has been used in each year for South and non-South regions.

REFERENCES

S. H. Altman, "Earnings, Unemployment, and the Supply of Enlisted Volunteers," *J. Hum. Resources*, winter 1969, *4*, 38–59.

——— and R. J. Barro, "Model of Officer Supply under Draft and No Draft Conditions," in *Studies Prepared for the President's Commission on an All-Volunteer Armed Force*, Vol. *1*, Washington 1970.

——— and A. E. Fechter, "The Supply of Military Personnel in the Absence of a Draft," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 19–31.

A. E. Fechter, "The Supply of First-Term Military Officers," Institute for Defense Analyses, Mar. 1967.

A. C. Fisher, "The Cost of the Draft and the Cost of Ending the Draft," *Amer. Econ. Rev.*, June 1969, *59*, 239–54.

———, "The Cost of Ending the Draft: Reply," *Amer. Econ. Rev.*, Dec. 1970, *60*, 979–83.

B. P. Klotz, "The Cost of Ending the Draft: Comment," *Amer. Econ. Rev.*, *60*, 970–78.

College Placement Council, *A Study of 1969–70 Beginning Offers by Business and Industry*, Bethlehem, Pa. 1970. Also for various years.

Report of the President's Commission on an All-Volunteer Armed Force, Washington 1970.

U.S. Congress, House, Committee on Armed Services, *Review of the Administration and Operation of the Selective Service System*, 89th Cong., 2d sess., June 1966, pp. 10042–43.

U.S. Office of Education, *Projections of Educational Statistics*, Washington 1969.

COMMUNICATIONS

The Market for a Semipublic Good: The Case of the *American Economic Review*

By YORAM BARZEL*

Operationally, the theory of public goods is now at a serious impasse.[1] The two optimum conditions—the total condition and the marginal condition—cannot be simultaneously applied in actual policy. The total condition, requiring that total benefits are not less than total costs, cannot be applied unless benefits can be evaluated. To appraise these, we have to know something about individual utilities or, in operational terms, we have to know something about demand for the public good. But if the marginal condition, equating price to marginal cost, is actually used, no market data is generated. This is because the marginal cost of producing an extra unit of service from an already produced public good is zero, and, in fact, we normally fail even to keep track of the quantity demanded at this price. So here we find ourselves in a position where the total condition cannot be evaluated. Alternatively we may forgo pricing at marginal cost and in this way obtain data on the demand for the public good at some positive price; but then the marginal condition is necessarily violated, and there is no a priori information to suggest that this is less wasteful than violating the total condition.[2] Buchanan's suggestion (see p. 194) to use the model of a perfectly discriminating monopolist to resolve the dilemma does not, as Samuelson has pointed out (see 1967, p. 199), bring any new operational elements into the picture.

I

In this paper an attempt is made to get around the dilemma by studying, not a purely collective good, but a semipublic good—*The American Economic Review* (*Review*). The commodity considered here, unlike a TV signal, is not strictly a public good. The ideas contained in a publication such as the *Review* constitute the core of its public-good component, since their availability to some does not reduce their availability to others. However, ideas are not directly "consumed." For the consumer to obtain access to ideas, they must be published and disseminated at costs which increase with the number of consumers. Casual empiricism suggests that the costs of printing and distributing such a publication are roughly proportional to circulation volume. On the other hand, the costs of editing and setting the publication in print may be an extension of the public-good aspect of this commodity where costs are independent of the volume of output. It thus seems proper to consider the *Review* as a semipublic good—a mixture of a purely public and purely private good. It is important to note at the outset that, while the above considerations are common to other kinds of publications, in this case the cost of the original research is borne not by the consumer but by scholars, universities, and foundations, etc. In this respect the learned journals differ from books and magazines priced so as to pay the authors' fees or royalties. Rather, ironically enough, they resemble commercial TV where the cost is not covered by the viewer but by the program benefactor, the advertiser.

---

* Visiting research fellow, University College, London, on leave from University of Washington. I wish to thank John E. Floyd and George Borts for their comments. Richard McDonald provided most able assistance.

[1] See in particular the exchanges (1964) between Jora Minasian and Paul Samuelson and between James Buchanan, Minasian, and Samuelson in 1967.

[2] The situation is analogous to the "uncertainty principle" in physics where, of two related variables, one can be measured only at the cost of foregoing measurement of the other.

It should also be noted that the *Review* is at least in part a producer rather than a consumer good. Consequently, the demand for it is a derived demand and the utility generated by it is "derived utility."[3] With this reservation in mind, it is appropriate to apply the analysis for collective consumption goods to this good as well.

In the literature on public goods, the discussion usually focuses on the public good itself, comparing the utility derived to cost. Here we will focus instead on the flow of services generated by the public, or the semipublic, good. The two approaches should, of course, produce identical results, and the choice between them is simply a matter of convenience. The service flow approach is more convenient for the present purpose.

In exploring the relation between pure public goods and semipublic goods, concentration upon the semipublic good might at first appear to be an undesirable compromise. However, in reality the conceptual difference between the two is minor. We can write a simplified version of the total cost function for the flow of services from the good under consideration (a single issue of the *Review*) as $TC=aX+C$, where TC is total cost, X is the level of output, and a and C are constants. In the case of a purely public good, $C>0$ and $a=0$, while in the case of a semipublic good, $C>0$ and $a>0$ (for a purely private good $C=0$, $a>0$).[4] Notice, then, that for both the pure public good and the semipublic good, the average cost function declines throughout with the marginal cost necessarily below the average cost. Thus, once we take into account the presence of a positive marginal cost (the a term) in the semipublic good,[5] most of the propositions for a purely public good will hold for a semipublic good as well.

In terms of the information potentially available, however, the difference between the two is of considerable importance. Since the marginal cost is zero, if the purely public good is priced to equate price with marginal cost, no market information is generated. If, on the other hand, a semipublic good is priced at marginal cost, it may provide us with information that purely public goods cannot—market information on utility generated. In other words, due to its purely private component, a semipublic good affords information on quantities and prices, and if this information is sufficiently varied, we may be able to obtain an approximation of its demand function.

Furthermore, if one tries to give examples of purely public goods, the list is almost exhausted after national defense, TV signals, and lighthouses. On the other hand, it is easy to compose a long list of semipublic commodities, to include books, magazines, newspapers (in fact everything printed), records, tapes, electricity where the power plant is not used to capacity, and so on. Since the policy problems raised by semipublic goods are of the same nature as those for purely public goods, obtaining information on the subject is important on its own.

II

The main objectives of the empirical work are to obtain estimates of the cost of producing the *Review* and of the benefits derived from it, to determine whether the optimum conditions of producing such a good are met and, if not, what may be done to meet them.

In the empirical work, data problems often prevent correct measurement of some of the desired quantities. On the demand side, observed price is not the subscription rate for the *Review* but membership dues to the American Economic Association. For some but not all years the inside cover of the *Review* states what portion of the dues applies to subscription. It is not possible, however, to buy the subscription at that price, so this information is neither complete nor operational. We have to view the demand, then, as comprising both the *Review* itself and other services provided to members by the association, such as organizing of meetings, supplying of directories and special supplements, and vesting of prestige in its members. To the extent that the level of such extra services did not vary much over the period studied, the misspecification of the demand function on this account may be ignored. It will be shown below that the actual level of these services did in fact remain quite stable. An additional service provided by the Association since 1963 is in publishing a second journal, the *Journal of Economic Abstracts* succeeded later by the *Journal of Economic Literature*. To avoid the effects on price and cost of these additional publications, the empirical work is terminated at 1965, the last year before the price of the *Review* was increased to meet the additional cost of the new publication. We will see later what can be said regarding these additional publications.[6]

---

[3] Richard Schmalensee has recently shown that under monopoly conditions, the observed increase in consumers' surplus due to a fall in price of a factor of production may be an underestimate of the true change in welfare.

[4] If "all enjoy in common" a collective good, the case where $a=0$ characterizes a pure collective good, since no additional cost is necessary for an additional consumer to enjoy it. If $C=0$ we are at the other extreme where there is no element of communality. We occupy the in-between situation of a semipublic good if $C>0$, constituting the common component of the good, and $a>0$, constituting its private component. (A broader definition of a semipublic good is where the marginal cost is positive but less than average cost.)

[5] This may be done by netting the demand curve from these costs. The resulting relationship between quantity and price may be thought of as a demand curve for a purely public good.

On the cost side we encounter a different problem that is probably much more significant; this is that two major cost elements associated with the *Review* are not borne by it: the cost of doing the original research (i.e., writing the "editorial copy") and the cost of refereeing. Who is bearing these costs? The latter seem to be borne entirely by the referees. However, they presumably get adequate psychic compensation from performing this function, since they are doing it voluntarily, and there is no need to view this aspect of publication as involving any net social cost.

The problem of the cost of research going into the *Review* cannot be dismissed so easily. True, part of it may be a consumption good to those engaged in it; for this part, as for refereeing, there is no problem. But, quite clearly, a substantial portion of the research would not have been forthcoming had it not been paid for either by universities, directly in support or indirectly in salary increases, or by other institutions with the objective of promoting knowledge. The institutions providing the funds almost never benefit directly from the research. Rather, their support for research is a recognition of the public good aspect of research and of the need for nonmarket sources of funds to sustain it.

The direct users and beneficiaries of the research results are the readers of the *Review*. It is thus appropriate to ask, What are the total consumers' benefits to readers of the *Review* and are these large enough to justify the use of resources devoted to research and to its publication? To the extent that the publication component of costs is met by subscription fees the above question is reduced to a comparison between the consumers' surplus derived from the *Review* and the cost of research. While data on the cost of research going into the *Review* are not available, an estimate of the benefits may be used to determine what research costs can be justified by demand considerations.

In the empirical work, then, we will on the one hand, attempt to establish the demand curve for the *Review*. On the other, we will try to obtain an estimate of the cost function of publishing the *Review* (but not of conducting the original research). First, we want to establish whether the casual notion of economies of scale (or its counterpart, declining average cost) is valid in regard to editing the *Review*; then we will try to determine the interactions of such economies with the change in demand patterns over time.

---

[6] It is sometimes claimed that since a single copy of the *Review* can be, and frequently is, read by many, this is another dimension of the public good aspect of the *Review* and the observed demand thereby understates the total demand. It should be recognized, however, that the owner of a copy of the *Review* has the option of excluding others from using it. If he does not do so, the demand of these other users is reflected in the utility and demand function of the owner of the copy. This is parallel to the utility from and market manifestation of the demand for guest-room furniture by the host rather than by the guest.

The American Economic Association publishes annually, in the *Papers and Proceedings* (*Proceedings*), a financial report as well as data on circulation and on the size of the journal in terms of the number of pages. The period covered in the regression analysis is 1930–65; consistent time-series are available only from 1930, and 1965 is the last year before fees were increased to cover the *Journal of Economic Abstracts*. Since the format of the journal changed little over that period there is no need to worry about this aspect of quality. On the other hand, the number of pages per issue shows a rising trend over time. This variable, then, should be controlled in the estimation process.

To account for economies of scale in production we consider the following:

total real cost $=f$ (number of pages, circulation).

The basic unit of output is the printed page, and output can be expanded either by an increase in circulation while holding the size of the *Review* constant, or alternatively by increasing the number of pages while maintaining the level of circulation. The estimation form used is:[7]

$$TC = a_0 + a_1P + a_2P^2 + a_3C + a_4C^2 + a_5PC + U$$

where TC is total real annual cost of publishing and printing, P is the annual number of pages,[8] C is average annual circulation,[9] and U is the random term. In the absence of economies or diseconomies of scale, only a_5 will differ from zero. If any of the other a's differs from zero, average cost will in general, differ from marginal cost.

The regression results (with t-ratios in brackets), using an alogarithm that maximizes R^2 ,are:

$$\begin{aligned} TC = \underset{(16.18)}{27{,}171} - \underset{(5.09)}{5.067C} + \underset{(3.33)}{.000104C^2} \\ + \underset{(3.71)}{.0041CP} \qquad R^2 = .976 \\ D\text{-}W = 2.18 \end{aligned}$$

The goodness of fit, the absence of serial correlation, and the sharpness of the coefficients are all gratifying. A closer examination of the results reveals some peculiarities at the lower range of the variables, probably reflecting the fact that during the early 1930's, a period during which circulation and the number of pages in the *Review* declined, total nominal costs remained fairly constant and total real costs increased.

Abstracting from these first few years, the results largely conform to the casual notions of an absence of economies of scale with respect to the size of the *Review*, and the presence of such economies to its circulation.[10] The latter, however, persisted only until circulation reached some 16,000 (around 1962), at which point the marginal cost of an extra copy became equal to the average cost. This finding is consistent with the actual experience of the AEA. Given the tendency of the Association for average cost pricing and the lag in price adjustments, as circulation grew prior to 1962, while average cost was falling, the Association found itself quite regularly in the "black." In more recent years, as circulation has continued to grow along with a rising average cost, deficits have become more common, eventually resulting in an increase in real price.

It is important to repeat that the cost of the original research is entirely independent of circulation and of the size of the *Review*.

---

[7] I am indebted to Sanford Berg for drawing to my attention that the logarithmic form I used in an earlier draft is inconsistent with the theoretic formulation.

[8] TC and P are defined for the *Review* alone, to exclude the *Proceedings*, since the latter is not edited by the editorial staff of the *Review*. The deflator used for TC is the wholesale price index exclusive of food and farm products.

[9] Circulation is taken as the number of copies of the *Proceedings* issue printed during the given year; it is almost identical with the average number of copies of the *Review*. Circulation at 3,700 was lowest in 1934 and at 20,000 was highest in 1965. The annual number of pages ranges from 792 in 1932 to 1,327 in 1965.

[10] Regressions for the *Proceedings* issue where recorded costs are almost entirely for printing and circulation, are even clearer in showing no scale economies with respect to these activities.

TABLE 1—REGRESSION RESULTS FOR THE DEMAND FUNCTION*
(*t* ratios in brackets)

	Price	No. of *BA*'s	No. of Pages in *Review* Alone	No. of Pages in *Review* & *Proceedings*	R^2	*D-W*
1.	−.50 (3.02)	1.01 (10.20)		−.15 (.69)	.96	.25
2.	−.53 (3.23)	.73 (4.49)	.78 (1.42)		.96	.24

* All variables in logarithms.

Had this cost item been included in the total cost, its only effect would have been to correspondingly increase the constant term. This in turn would mean an increase in average cost with no effect on the marginal cost.

III

So much for the cost side; we will now turn to demand considerations. How should close substitutes to the *Review*, particularly other economic journals, be handled in this context? Since prices of close substitutes tend to move together over time, it is dificult statistically to separate the own elasticity from the cross-elasticity. If we allow only the own price to enter, we tend to get the industry demand elasticity—that facing the individual producer when all producers n the industry change prices together. But this is precisely what we would like to obtain, since the basic choice is between doing and not doing the research at all, rather than between doing the research for the *Review* versus doing it for some other journal. The problem is complicated, however, by the act that during the period considered the range of substitutes has broadened, with the nception of such journals as *Econometrica*, *The Review of Economic Studies*, and *The International Economic Review*. This, of course, has lowered the effective price of subtitutes but in a way extremely difficult to ake into account empirically. The trend in he real price of the *Review* over the period of tudy is basically downwards, the same diection of change as in the newly introduced ubstitutes; but, of course, the latter fell in price more than the *Review*. Subscribers, then, have had a greater incentive to switch to these substitutes than would have been the case if all the prices had changed by the same proportion. Consequently, not explicitly accounting for the new journals introduces a downward bias to the estimate of elasticity of the industry demand obtained by observing the *Review* alone. It is interesting to observe that these new journals themselves probably represent a response to the increase in market size and to the resultant sufficiently expanded demand for such highly specialized products.

With the above caveats, the general form of the demand function is:

$$q = f(\text{real price, real per capita income, population, number of pages})$$

For estimation, the form used is:

$$\log C = b_0 + b_1 \log Pr + b_2 \log Y + b_3 \log Pop + b_4 \log P + V$$

where C is circulation, P is the number of pages, and V is the random term. Pr is price, defined as the annual subscription rate deflated by the consumer price index, and Pop is population, alternatively defined as total labor force or as total number of individuals possessing bachelors' degrees.[11] It is not at all clear that per capita income is the relevant variable here, since the *Review* as mentioned above may be considered a producer rather than a consumer good. In the actual regressions this variable was never significant and, in fact, had a negative sign. Consequently, the results with this variable present are omitted from the findings presented in Table 1. As to the choice between the labor force or the number of *BA*'s, the latter performs slightly better. It is also, on a priori grounds, a better proxy for industry size, the results are reported only for the number of *BA*'s.

---

[11] Ideally, we would like to measure the size of the industry consuming the commodity under consideration. The best measure of the size of the relevant industry is membership in the American Economic Association. This however, may lead to somewhat spurious results. The number of *BA*'s was considered as a reasonable alternative. This was measured as a 25-year moving sum of *BA*'s granted up to the given year; for 1917–1957, where the number of new *BA*'s is not available for odd years, it was interpolated from the two adjacent even years.

The most impressive result is that of the price coefficient. It is quite stable from one regression to another (including those not reported here) and has a consistently small standard error. The size of the coefficient, indicating a demand elasticity of about 0.5, may appear low. But one should bear in mind another aspect of this commodity as a public good—that for the individual a second copy has very little, if any, use. Thus a fall in price may induce additional demand by new readers but virtually no additional demand by current readers.[12] The number of *BA*'s (as well as the labor force) is highly correlated with the number of pages, and it is hard to sort out their individual effects. In any case it seems that the combined number of pages in the *Review* and *Proceedings* is not a relevant variable. Therefore, regression 2 may be considered the better. In subsequent sections, I use the results of demand regression 2 dealing with the *Review* alone.

IV

Let us now turn to see how these results can be interpreted in reference to the public-good problem. Again, we seek answers to two basic problems. One is whether to engage in producing the good at all—the total condition. The other is what quantity of resources should be devoted to that good—the marginal condition. To answer the first we have to determine whether the area below the demand curve is at least as large as the total resource cost of producing the journal, where the alternative cost of the resources engaged in research for the *Review* is included in the cost. Here, however, we do not attempt to measure the resource cost of doing research. Instead we attempt to measure the benefits from such research, thereby obtaining a measure of the maximum amount of resources that should be devoted to that activity.

The area under the demand curve, taking the form of the demand equation and using the numerical results of regression 2, is given by

$$\int_{Pr}^{\infty} (\text{antilog} - 4{,}147)$$

$$(Pop)^{.729}(P)^{.781}(Pr)^{-.529}dPr$$

The area, of course, depends on the values taken by the three variables. However, before inserting such values, it should be recognized that it does not make much sense to pretend that we can safely extrapolate the demand curve to an infinite price.[13] On the other hand, if we restrict ourselves to the observed range of prices, we obtain a highly conservative estimate of the area under the demand curve. The highest real price observed is that for 1941, when the annual subscription rate was $5.00 and the consumer price index stood at 51.3 (1957–59 = 100). The lowest real price observed is for 1961, ($6.00 deflated by 104.2). The 1941 real price is then higher than that of 1961 by some 80 percent. Using this range of prices and the 1961 levels of the number of *BA*'s (6,371,000) and number of pages (1168), the value of the above integral, in regression units, is 24.55 − 19.16 = 5.39. Since the price index is 100 for 1957–1959 and the quantity is measured in units of a hundred, when adjusted the area becomes $53,900 (in 1957–59 prices).[14] This represents the extent by which 1961 subscribers to the *Review* benefited by not having to pay the 1941 price.

In diagrammatic terms, this is shown in Figure 1, where D_{61} reflects demand in 1961 and the intersection point of this curve with

---

[12] On the other hand, consumers are limited in terms of buying less than one whole unit. So a fall in price normally will lead to shifts from buying zero units to buying one unit.

[13] In this case the area is infinite, so we can disregard this as an economic impossibility. (Or are we experiencing Nirvana by merely having the *Review*?)

[14] The value of $53,700 is a point estimate and subject to a sampling error. It should be viewed, then, only as pointing to orders of magnitudes involved. The same applies to all subsequent estimates. Also, it contains an element of upward bias due to the above mentioned bias in estimating the demand elasticity.

AC reflects that fact that the *Review* normally covers its costs through subscription fees. The portion under the curve bounded by P_{41} and Q_{61} is a measure of total benefits from the *Review* under the assumption that the demand curve has an elasticity of .53 up to the 1941 price and then becomes perfectly elastic. The area $P_{61}xQ_{61}$ represents actual *Review* office and publication costs, and the area bounded by P_{41}, P_{61}, and D_{61} is the "consumers' surplus" evaluated above at $53,900.

The value of $53,900 should not be entirely attributed to the fall in price from 1941 to 1961. The values of the other independent variables also changed, and it is of interest to try to allocate the $53,900 among these factors. In 1941 the *Review* had some 20 percent fewer pages than in 1961, and the estimated number of *BA*'s was less than half of the 1961 level. Thus, the shift in quantity demanded is due not only to a move along the demand curve but to a shift of the curve itself. Again, when we now compare the area below the curve in 1941 and in 1961, we ignore the "upper tail" of the demand curve and the change in the upper tail between the two periods. When the area under the curve due to shifts in demand is calculated we obtain $\$10{,}000 \times \{(25.55-19.16)-(10.90-8.51)\} = \$30{,}000$. So, out of the $53,900 area below the 1961 demand curve bounded by the 1961 and 1941 prices, $30,000 may be attributed to a shift in demand and the remainder, $23,900 to a shift along the demand curve.[15] In terms of Figure 1, the $30,000 is the area bounded by the two prices and lying between the two demand curves, and the area to the left of D_{41} is measured as $23,900. The former is a measure of the added consumer benefits due to the increase in number of consumers and of the improvement in the quality of the product (in terms of number of pages). The latter is a measure of the realization of economies of scale, resulting in a fall in real price of the *Review*.

The $30,000 might be further broken down into two components: the number of *BA*'s and the number of pages. They are, respectively, $21,900 and $8,100. The shift in demand producing the $21,900 figure is basically due to an increase in the size of the economy (in terms of population and in terms of per capita income, both reflected here by the number of *BA*'s). As the economy expands, demand curves tend to shift outward, and the value of a public good grows correspondingly. Since the cost of the public good is independent of the intensity of its use, the difference between its value and its cost grows with the economy's size, thereby increasing "productivity" in the economy. In fact we may want to consider the entire $30,000—or even the $53,900—figure as the appropriate one, since both the increase in number of pages and the fall in price can be viewed as a response to the increase in circulation.

If we assume that the *Review* itself typically breaks even in its financial operations, and if we further assume (without much conviction) that the total condition was just met in 1941, the annual amount of resources devoted to research for the *Review* could have increased from the 1941 level by as much as $53,900 (in 1957–1959 prices) without violating the total condition. This, of course, still fails to account for the change in the upper tail of the curve.

A mild attempt to account for this upper tail is to assume that the demand function retains its form for a certain price range but above that point becomes perfectly elastic. If we assume that the price at the upper end of the range is twice that of the 1941 real price, the consumers' surplus increases from $53,900 to $149,100. If we allow the price at the upper end to be three times the 1941 level, the area increases further to $220,400 (see Figure 1).

When for purposes of comparison with 1941 we use 1965 (the last year included in the study) rather than 1961, we find the number of *BA*'s and the number of pages have increased but the price of the *Review* has also gone up. The 1965 consumers' surplus arising from not having to pay the 1941 price is $39,000. Now if the price at the upper end of the range is twice the 1941 price we get $158,600, and it if is three times the 1941 price we get $248,500. All these figures represent an attempt to approximate the total condition. They reflect, of course, only the benefit side of the picture, against which we may match total cost.

---

[15] Such allocation necessarily contains an arbitrary element.

Let us now consider the marginal condition. For a commodity such as the *Review* or, for that matter, many other public goods, the units in which it is desirable to count output are far from obvious. Output can be expanded by printing more copies of the *Review* without changing its contents, or output can be expanded by devoting more resources to improve the quality of the papers going into the *Review.* Also the size of the *Review* can be increased by either adding more issues per year or by increasing the number of pages per issue.

The last alternative seems to be most reasonable to investigate further. This is one of the ways in which the *Review* has expanded in the past, and it is relatively easy to control. Thus, let us investigate the effect which a 1 percent increase in the size of the *Review* may have on total cost and on total benefits. Given the 1965 conditions, this is equivalent to about thirteen extra pages, the length of a typical article. Without a change in price, this would imply a shift in the demand curve to the right by .776 of 1 percent (the regression coefficient above for number of pages). This means an increase of $300 in consumers' surplus when the demand curve is bound by the 1941 price,[16] $1,200 if bound by twice that price, and $1,900 if bound by three times that price. In addition, the number of copies demanded increases by .78 of 1 percent and at the given price of the *Review,* revenues increase by the same proportion. Circulation in 1965 was 20,000 and the subscription rate $8.00. The increase in revenue, then, will be about $1,100 (in 1957–1969 prices). This addition to revenues, however, will be more than matched by the added expenses of $1,400 due to increased circulation cost and additional editing and publishing of the extra pages. Apart from this $300 excess of cost over revenue, the question boils down to determining whether the total research cost required to generate one more article[17] of the same average quality as the other *Review* articles is less than $300, between $300 and $1,900, or more than $1,900. The first figure provides a lower bound as to the value of the article; the middle range still probably does not overestimate its value; but if total cost of an article exceeds $1,900, there may be serious doubts as to whether the benefits generated by the article are sufficient to cover the added cost.

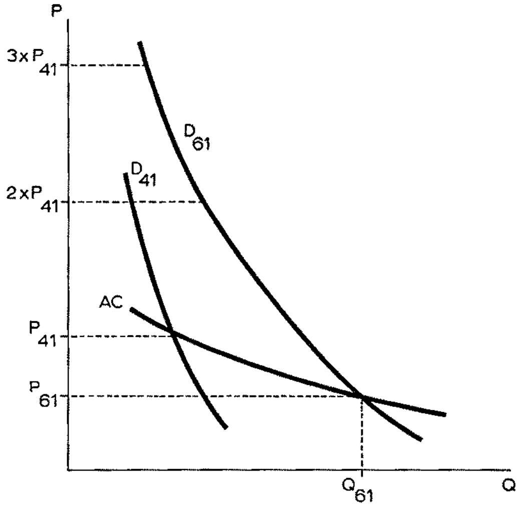


FIGURE 1

An important qualification should now be made here to the way in which the marginal condition was implicitly applied above, with regard to the timing of research and publication. The decision rule which should be applied is the maximization of the present value of the research investment. The research for a given journal article may be undertaken at different points in time. Since publishing research at a given date excludes it from being published at any other date, the article may be considered as mutually exclusive with itself over time. If demand is growing steadily over time, as is the case for the *Review*, maximizing the present value of an article necessarily leads to a positive present value. Consequently the break-even point in terms of the total condition—i.e., when net present value is equal to zero—is not an optimal point: the activity has to be postponed until the appropriate marginal condition is satisfied, when the time rate of change in benefits is equal to that in costs.[18] In fact, under the stated conditions the marginal condition is the necessary and sufficient condition for optimum.

---

[16] The 1965 real price, however, is only about 35 percent higher than the 1941 real price. With the 1941 price constituting an upper bound on demand, consumers' surplus is necessarily low.

[17] Only a fraction of the articles written are published, and the cost of the unpublished ones (the "dry holes") should be counted as part of the total cost.

V

In discussing the demand for the *Review*, it was mentioned that subscribers receive benefits from the Association in addition to the *Review* itself. We propose to measure the value of these services to subscribers as the difference between 1) the real subscription fees and 2) the costs associated with the *Review*, defined as predicted (real) costs from the above cost equation. For all but one year, the difference, in 1957–59 values, varied between $2.50 and $4.30 with no apparent trend. The changes from year to year mostly reflect the drift in costs and in real fees on the one hand and in the discrete changes in fees on the other. We may thus conclude that while demand by subscribers is for both the *Review* and the other services provided by the Association, disregarding the latter in the demand equation does not lead to a serious distortion of the results.

Since 1963, the AEA has been publishing a second journal: the *Journal of Economic Abstracts* in the years 1963–68 and the *Journal of Economic Literature* (*JEL*) since 1968. How does this fact relate to the rest of the study? From 1963 to 1965, the *Journal of Economic Abstracts* was priced separately from the *Review* and did not affect either the subscription fees or the cost of the *Review*. Since 1969 the format of the latter has been drastically changed, and at the same time, the *Journal of Economic Abstracts* was superseded by the *JEL*. It is not possible to separate the simultaneous effects of the two changes.

Over the 1966–68 period, however, the format of the *Review* was the same as in the previous years. Given the subscription rate and the number of *BA*'s for these years, it is possible to predict what would have been the quantity demanded (i.e., circulation of the *Review*) in the absence of the new journal. The difference between the actual number of subscribers and that predicted can be attributed to the introduction of the *Journal of Economic Abstracts*. However, given the positive serial correlation in the demand regression, we should expect the residuals for 1966–68 to be at the same level as for the immediately preceding years. The residuals for 1963–65 averaged approximately .07, indicating a positive deviation of about 17.5 percent. The average difference between (the logarithms of) the observed and predicted subscription for 1966–68 is about .125, or a positive deviation of about 33 percent. We thus conclude that subscription to the *Review* increased by about 13.5 percent (antilog of $.125-.07=.055$) in response to the introduction of the *Journal of Economic Abstracts*.

VI

The main purpose of this paper has been to obtain some empirical evidence on the demand for a semipublic good, the *Review*, and thereby to demonstrate that at least for semipublic goods there exists a partial solution to the operational dilemma in which the theory of public goods is mired. It is true that the empirical analysis is incomplete, due to lack of data on the cost side on the part of contributors to the *Review*.

This, however, should not be a serious shortcoming in terms of assessing the degree to which the optimum conditions are met for other collective goods, since for many of these the required cost data are relatively easy to obtain. It should be pointed out in closing that even for a commodity such as a TV signal, some usable market data are available to estimate demand. Expenditures

---

[13] For a detailed discussion of this point, see Stephen Marglin and my article.

on TV sets reflect demand for programs. Expenditures on individual and centralized antennas also reflect demand and sometimes can be used to determine the demand for a specific extra channel. Such information obviously cannot provide accurate estimates of the demand for TV signals, even less of that for specific programs. Nevertheless, given the current state of knowledge as to the demand for TV signals, analysis of the available data may be valuable in making the theory of public goods operational.

REFERENCES

Y. Barzel, "The Optimal Timing of Innovations," *Rev. Econ. Statist.*, Aug. 1968, *50*, 548–55.

J. M. Buchanan, "Public Goods in Theory and Practice: A Note on the Minasian-Samuelson Discussion," *J. Law Econ.*, Oct. 1967, *10*, 193–97.

S. A. Marglin, *Approaches to Dynamic Investment*, Amsterdam 1963.

J. R. Minasian, "Television Pricing and the Theory of Public Goods," *J. Law Econ.*, Oct. 1964, *7*, 71–80.

———, "Public Goods in Theory and Practice Revisited," *J. Law Econ.*, Oct. 1967, *10*, 205–07.

P. A. Samuelson, "Public Goods and Subscription TV: Correction of the Record," *J. Law Econ.*, Oct. 1964, *7*, 81–83.

———, "Pitfalls in the Analysis of Public Goods," *J. Law Econ.*, Oct. 1967, *10*, 199–204.

R. Schmalensee, "Consumer's Surplus and Producer's Goods," *Amer. Econ. Rev.*, Sept. 1971, *61*, 682–87.

The Neo-Factor Proportions Theory of International Trade: An Empirical Investigation

By THOMAS C. LOWINGER*

Traditionally, the factor proportions theorem viewed the pattern of international trade as the outcome of the interaction between the production characteristics of traded commodities (with respect to capital-labor ratios) and the relative endowments of countries with regard to labor and capital. Given similar conditions of production in two countries, the prediction was that each country will export the commodity which, in the absence of trade, is relatively intensive in the country's abundant factor.

The controversy, that arose subsequent to the publication of the Leontief Paradox, had the laudatory effect of directing some effort toward the formulation and testing of alternative theories of trade structure. One especially promising avenue was opened as a result of the work on human capital carried out by Gary Becker and Theodore Schultz.

The "human capital" approach departed from the Heckscher-Ohlin-Samuelson formulation of the model which identified "capital" with physical plant and equipment and "labor" with counting bodies regardless of skill.

The work done by Donald Keesing and Helen Waehrer dealt with the effect of labor skills or human capital on the structure of trade. Specifically, they suggested that the category of factors aggregated under the catchall term labor should be subdivided based on their skill content (i.e., reflecting varying intensities of human capital). Predictions about trade patterns could then be made on the basis of the relative skill endowments of the trading countries.[1]

If one views capital as an activity giving rise to a flow of income (or product) over time, it becomes evident that there will be some overlap between commodities which have high components of "physical" capital and human capital. This provides a priori justification for combining the contributions of human and physical capital to the production process into a single measure of man-made resources. Hal B. Lary suggested that value added per employee can be used as a proxy for the industry's "overall" factor intensity (of human and physical capital). Since both the acquisition of tangible capital and the upgrading of skills involve acts of investment (or saving), such an approach commands a great deal of intuitive appeal.

This paper is limited to the study of the structure of trade conducted between a semi-industrialized country and a group of countries possessing an advanced industrial base.[2] To analyze the commodity composition of trade between a semi-industrialized country that has a relatively small and unsophisticated manufacturing sector and the industrialized countries of Western Europe and North America, the neo-factor proportions theory is evoked.

---

*Assistant professor of economics, Washington State University. The constructive comments of John T. Donnelly and a referee of this *Review* are gratefully acknowledged.

[1] Keesing's results suggest that the United States, being relatively well endowed in skilled labor, tends to export those goods which embody in their production a high proportion of skilled labor (its abundant factor). More recently a study by Gary Hufbauer related the skill endowments of twenty-four countries to the skill content of their commodity trade. Notwithstanding the imperfections inherent in such international comparisons, the study obtained high Spearman Rank correlation coefficients between national skill endowments and skill characteristics embodied in traded manufactured goods.

[2] In explaining trade patterns among industrial countries, it is appropriate to consider not only the factor proportions theory, but also what has been referred to as the "neo-technological" theory. The analysis of technological change as a source of trade has involved such disparate hypotheses as the "product cycle," the technological gap, and the role of economies of scale in modern industry. See Harry G. Johnson, pp. 15–16.

I. Brazil's Factor Endowments Relative to Her Main Trading Partners

The factor proportions account views trade patterns as the outcome of the interaction between the production characteristics of commodities (with respect to their "factor" intensities) and the countries' relative availability of these factors.

This section inquires into Brazil's factor endowments relative to her main trading partners, the industrial countries of Western Europe and North America.

Human Capital

There are various ways of comparing human resource levels internationally. In Table 1, fourteen countries at varying stages of development are compared based on three measures. In conjunction, these indexes could be used to represent the countries' stocks of human capital. The evidence presented in Table 1 indicates that based on the Harbison index, Brazil ranks lowest in its endowment of human capital.

Based on the ILO classification system, another comparison was attempted between the "skill composition" of Brazil's (and other Latin American countries') labor forces and those of several industrial countries. Table 1 presents two ratios as estimates of the availability of skilled manpower.

Ratio A refers to the proportion of the economically active population that is included in ILO category 0 (professional, technical, and related workers). Ratio B includes, in addition to category 0, workers in occupational category 1 (administrative, executive, and managerial personnel). Based on either ratio A or B, it is evident that Brazil had an overall shortage of skilled labor compared to North America and several Western European countries. For example, based on ratio A, the labor forces of the United States and Canada contain on the average three times the skilled workers of the Brazilian labor force.

Additional evidence presented elsewhere substantiates the existence of a serious shortage of Brazilian skilled personnel at all levels and especially of engineers, technicians, and managerial personnel.[3] Given the inadequacy of educational opportunities in Brazil and the shortage of facilities for the training of skilled workers, the skill composition of the country's labor force is likely to change very slowly in the near future. Thus, relative to the industrial countries, the position of Brazil as a skill-deficient country will be maintained in the near future.

TABLE 1—OCCUPATIONAL CHARACTERISTICS OF THE LABOR FORCE AND THE LEVEL OF HUMAN RESOURCES FOR BRAZIL AND SELECTED COUNTRIES

	Year of Survey	Skilled Employees as a Proportion of the Economically Active Population[a]		Human Resource Level
		Ratio A	Ratio B	Harbison Index[b]
I. North America and Western Europe				
United States	1960	.108	.187	261.3
Canada	1961	.106	.183	101.6
Belgium	1961	.080	.106	123.6
United Kingdom	1961	.086	.112	121.6
West Germany	1961	.076	.107	85.8
Italy	1965	.053	.133	56.8
Netherlands	1960	.092	.123	133.7
Sweden	1960	.129	.150	79.2
II. Latin America				
Mexico	1960	.036	.044	33.0
Chile	1960	.049	.068	51.2
Ecuador	1962	.033	.036	24.4
Paraguay	1962	.032	.040	22.7
Peru	1961	.033	.048	30.2
Brazil	1963	.037	.048	20.9

Sources: Harbison and Myers, pp. 31–33; ILO, *Yearbook of Labour Statistics 1966*.

[a] Ratio A is the proportion of skilled workers in category 0 (professional, technical, and related workers) as of total employees. Ratio B is the proportion of category 0 and 1 (administrative, executive, and managerial) as of total employees.

[b] The Harbison index is calculated as a sum of two ratios: 1) students currently enrolled in secondary schools as a percentage of school-age population; and 2) students enrolled in institutions of higher education (colleges) as a percentage of school-age population multiplied by a factor of five.

Physical Capital

Reliable international comparisons of capital stocks are largely unavailable. Hufbauer estimated the capital stock per manufacturing employee of twenty-two countries. His procedure involved adding up the current expenditures for manufacturing investment be-

---

[3] Nathaniel Leff (1968b, p. 70) found that in the Brazilian capital goods industry the ratio of skilled to unskilled worker's wages is from 30 to 90 percent *higher* than a similar ratio prevailing in several industrial countries.

TABLE 2—ESTIMATION OF THE CAPITAL STOCK AND GROSS DOMESTIC PRODUCT PER CAPITA IN INDUSTRIAL COUNTRIES AND BRAZIL, 1964
(in dollars)

	Fixed Capital Per Employee in Manufacturing	*GDP* Per Capita
Industrial Countries:		
Canada	8,850	2,002
United States	7,950	3,014
France	4,900	1,534
Belgium	4,400	1,458
Germany	4,250	1,541
Netherlands	4,750	1,264
Sweden	5,400	2,032
United Kingdom	4,000	1,480
Italy	2,100	916
Brazil	2,100	193

Note: It was estimated that the share of manufacturing industries in gross capital formation was about 20 percent.

Source: For industrial countries: Hufbauer, Table 6, p. 77; for Brazil: Data on gross fixed investment taken from U.N., *Yearbook of National Accounts Statistics 1966.* Exchange rates employed are from: I.M.F., *Annual Reports On Exchange Restriction.*

tween 1953 and 1964 and then dividing by total manufacturing employment in 1964. No account was taken of inflation or depreciation of the capital stock.

Using a modified version of Hufbauer's procedure, Brazil's capital stock was estimated.[4] The results presented in Table 2 would qualify Brazil as a country scarce in physical capital relative to North America and Western Europe. However, one must stress the rough nature of both Hufbauer's estimates and my own.

Brazil's shortage of physical capital is probably not as acute as the shortage of certain skills in her labor force. In the first place, the flow of capital internationally is considerably less inhibited than the movement of labor across national boundaries. For instance, there was a considerable flow of capital into Brazil in the form of direct investments.[5] Secondly, given the wage differential that exists between Brazil and the United States, the net flow of human capital (in the form of skilled personnel) acts to *worsen* the initial discrepancy in skill availability between the two countries.[6]

Thus, it has been established that Brazil is seriously deficient in skilled labor in relation to her chief trading partners. Furthermore, the country is capital (both human and physical) poor relative to the industrial nations of North America and Western Europe.

II. Empirical Test of the Neo-Factor Proportions Explanation of Brazil's Trade

In this section we will test two "orthodox" explanations of the commodity composition of trade. The first is the simple factor proportions account, which uses differences in physical capital intensity as the primary explanation of the structure of a country's commodity trade. Second is the human capital or skill-intensity explanation of trade. In certain instances, the human and physical capital intensities have been combined into one "overall" measure of man-made resources.

Based on two-digit SITC categories, we obtained the proportion of Brazil's imports and exports of manufacturers that originated in or were destined for the markets of Western Europe and North America.[7] These ratios were then related to the characteristics of the traded commodities, with respect to capital per man, skill ratios, and wages per man.

---

[4] The main difference between Hufbauer's calculation and mine involves the use of exchange rates. Hufbauer used mid-period (i.e., generally 1958) exchange rates to convert local currency units into *U.S.* dollars. My procedure used the average of import category exchange rates (since multiple exchange rates were in effect in Brazil) to convert the annual data on Brazilian capital formation. This procedure is better suited to take account of the rapidly depreciating Brazilian currency resulting from the domestic inflation.

[5] Between 1955 and 1961 the flow of direct foreign investment into Brazil amounted to $714 million. See Leff (1968a, pp. 59–76).

[6] Further comparison that would reflect the rates of return of human relative to physical capital in Brazil and Western Europe (or the United States) should be useful in this connection. Martin Carnoy found the rates of return to schooling in Mexico, Chile, and Colombia to be significantly higher than the rates of return to physical capital. Unfortunately, no such comparison is currently available for Brazil.

[7] The industrial countries proportion (P_i) of Brazil's total trade was calculated based on the following formula:

When interpreting the results reported in Table 3, one must remember that virtually all of Brazil's imports of manufactured goods originated in industrial countries, but that no such clear-cut division exists in the destination of her manufactures exports.

When imports were ranked based on their human and physical capital intensity, the neo-factor proportions account fared well. These results support the hypothesis that Brazil's imports of manufactures included a high proportion of commodities that are intensive in capital—conceived in a broad sense to include not only the material means of production but also human skills. Unfortunately, this test cannot be used to discriminate between the physical and human capital explanations of the structure of Brazil's trade.

Furthermore, it is evident that the human and physical capital intensities of traded commodities did not provide a satisfactory explanation of Brazil's pattern of manufactured exports. (In two cases, the coefficients are negative as one would expect, but insignificant at the customary level.)

This latter outcome is not entirely surprising given that Brazil regarded the exports of manufactures as an incidental *by-product* of the evolution of her manufacturing sector, through the process of import substitution. The expansion of her industrial base in the post-World War II period was stimulated by means of various incentives and subsidies and in particular, the exchange and tariff privileges granted industries during the 1950's. Given the importance of domestic market size in the process of industrialization through import substitution, Brazil considered exports of manufactures as an outlet for the output that could not be accommodated domestically. Investment in local production was undertaken without regard to the future feasibility of exporting these products and with the knowledge that the domestic market was "secured" through protective policies adopted by the government.[8] Thus, Brazil's "exportable surplus" approach to International Trade appears to be largely responsible for the absence of a clear-cut pattern in her exports of manufactures.[9]

TABLE 3—RANK CORRELATION TEST OF BRAZIL'S COMMODITY TRADE CHARACTERISTICS AND THE DIRECTION OF HER TRADE

	Industrial countries' proportion in Brazil's imports of manufactures (1966)	Industrial countries' proportion of Brazil's total exports of manufactures (1966)
Capital per man	.450** ($t=2.523$)	−.028 ($t=.143$)
Skill ratio	.500* ($t=2.883$)	−.119† ($t=.600$)
Wages per man	.575* ($t=3.514$)	.127 ($t=.022$)

Note: * Significant at the 1 percent level; ** Significant at the 2 percent level; † Significant at the 30 percent level.

Source: Data on commodity characteristics were taken from Hufbauer, Table 3, pp. 59–70; Brazilian trade data are from U.N., *Commodity Trade Statistics 1966.*

$$P_i = \frac{C_i + V_i + W_i}{Q}$$

where:

C_i = dollar shipments from (to) Brazil to (from) Canada of the ith two-digit commodity.

V_i = dollar shipments from (to) Brazil to (from) the United States of the ith two-digit commodity.

W_i = dollar shipments from (to) Brazil to (from) Western Europe of the ith commodity.

Q = Brazil's total imports (or exports) of manufactured goods in SITC categories 5, 6, 7, and 8.

Thus, P_i gives us the proportion of industrial countries' trade in manufactured goods with Brazil in each two-digit SITC category.

[8] Joel Bergsman noted that between 1953 and 1959 there existed considerable uncertainty (and fluctuation) with respect to the real export rate. This, he believes, had a significant effect in discouraging non-primary products' exports and contributed to the inward-looking attitude of the Brazilian manufacturing sector (see pp. 35–36).

[9] Based on this approach a country exports only the surplus that is left over after the domestic market has

Given these inconclusive results, a different set of tests has been designed to analyze the competitive performance of Brazil's industries in international trade. Also, a different set of industry characteristics has been devised.

Numerous measures have been suggested in order to gauge the international competitiveness of an industry (e.g., the time trends of exports and imports or comparative production costs). In this study, one measure used is the trade balance (i.e., imports less exports of two-digit SIC industries) as a ratio of the value of industry's domestic sales. Presumably, those industries in which Brazil had *no* comparative advantage would rank high based on this index of competitiveness and vice versa. As a complementary indicator, we computed the ratio of "gross" imports to sales for each industry. Both ratios are based on the ability of Brazil's industries to compete in foreign markets and on the extent of penetration of industrial countries' exports into the domestic market.

With regard to the industries' factor intensities, the following variables were used: Value added per employee, two measures of the skill content of each industry, and a proxy for the capital-labor ratio. Following Lary, value added per employee has been used as a measure of the flow of human and physical capital into the manufacturing process. Using the flows of man-made resources into the industries' output, value added per employee is taken as an indicator of the overall capital intensity of an industry.[10]

Investment in human beings has been often referred to by the catchall phrase "training." It includes formal and technical education as well as on-the-job training. The individual who pays for some or all of the costs involved in improving his skill would have increased his lifetime earnings stream more than the person with less training. Assuming that workers in "skilled" occupations incur greater costs of investment in themselves than "unskilled" workers, the skilled workers should receive greater absolute wages than those received by the unskilled workers.

To test whether there exists a relationship between the interindustry skill mix and the industries' competitive performance, a measure of the industries' skill composition was constructed. Mrs. Waehrer constructed such a measure of industry's skill in her study of the wage structure of export and import-competing industries in the United States. The Waehrer index regrettably suffers from certain limitations as a measure of skill or as a proxy for human capital.[11]

The Brazilian Census classifies employees in each industry according to their actual earnings. Based on this data, we determine the number of skilled workers, using CR 15,000 per month as the cutoff point. That is employees earning above this salary would be classified as skilled and those falling below this as unskilled. The members of the following occupations were consequently classified as skilled: Engineers, managers and accountants, salesmen, billing supervisors, and foremen. The skill-index therefore measures the percentage of employees in each industry that earned in excess of CR 15,000 per

---

been adequately supplied. The idea of concentrating output in specialized production for the export market is shunned. These ideas are explored and documented for Brazil in Leff (1967).

[10] Robert Baldwin in a recent study combined data on physical capital with estimates of the cost of education as a measure of the human and physical capital intensity of traded goods. He noted that such a procedure may not be appropriate for a less developed country.

[11] The "occupational index" was defined as the percentages of employees in each industry, out of all employees, falling into six occupational groups designated as skilled. Consequently, the index will vary from industry to industry only with differences in the *distribution* of employees defined by her as skilled. This will conceal wide variations in skill content within each of the broad categories of skilled workers. As an illustration suppose that two industries rank equally in terms of Waehrer's skill index; but one of the industries has a higher proportion of managers and engineers, while the second industry has an equally large proportion of sales and clerical personnel. Clearly, from the standpoint of human capital intensity the first industry should rank higher than the second; however, this will not be reflected in the Waehrer skill index, because of her definitions of skilled workers. For details, see Waehrer, pp. 92–97.

month, or fell into the occupational categories listed above.[12]

The final variable attempts to measure the physical stock of capital per employee in each industry. Aware of the difficulties encountered in estimating the stock of capital, I employed as a proxy for this variable the horsepower per employee in each industry. As a test of the extent of association between the industry characteristics and the trade performance ratios, the Spearman Rank correlation test was used.

The results in Table 4 lend some support to the human capital variant of the neofactor proportions account.

It has been proposed here that the disparities in the availability of factors between Brazil and her main trading partners are likely to be most pronounced with respect to human capital. Consequently, Brazil would compensate for its scarcest factor by importing from Western Europe and the United States those commodities that embody in their production a disproportionate amount of human capital. These results are all the more convincing because they are based on several different ratios of an inindustry's trade performance, i.e., the industry's "gross" imports and its trade balance.

Thus, the comparative *dis*advantage of Brazil is most pronounced in the manufactured commodities requiring a great deal of skilled labor in their production. Physical capital intensity and the overall capital intensity did not perform as well in "explaining" Brazil's trade pattern.[13]

TABLE 4—SPEARMAN RANK CORRELATION COEFFICIENTS OF BRAZIL'S TRADE PERFORMANCE (1966) AND INDUSTRY CHARACTERISTICS (1959)

	Industry Characteristics[a]		
	Skill Index	Horsepower per Employee	Value Added per Employee
Ratio of industry's net imports to value of sales	.695* (t=4.103)	.077 (t= .329)	.328 (t=1.472)
Ratio of industry's gross imports to value of sales	.653* (t=3.658)	.339 (t=1.528)	.417+ (t=1.947)
Ratio of industry's imports to total imports	.533** (t=2.669)	.458*** (t=2.186)	.395+ (t=1.822)

Sources: Brazil's Industry Data are from IBGE, *Censo Industrial de 1960, Brazil*, VII; Brazil's Trade Data are from U.N., *Commodity Trade Statistics 1966.*

Notes: [a] Each industry's skill index is computed relative to all manufacturing industries; sample of 20 industries used.

Skill Index: The percentage of employees in each industry who earned in excess of CR 15,000 per month.

* Statistically significant at the 1 percent level; ** Statistically significant at the 2 percent level; *** Statistically significant at the 5 percent level; + Statistically significant at the 10 percent level.

These results are not without an intuitive appeal. Differences in the skill characteristics of labor among less developed and industrial countries are very pronounced and may be getting even more unequal over time. Basically, skilled people are needed to train other skilled people and the disparity may be further widened because of the tendency of highly skilled members of the labor force to migrate to countries enjoying a higher standard of living, especially the United States.

In 1969 Brazil exported about $250 million worth of manufactured goods or about 11 percent of her total exports; at the beginning of the 1960's exports of manufactures were only about $25 million per annum. Yet this change is not as encouraging as it may appear because it does not represent a fundamental departure from the post-World War II pattern of industrial growth based on import-substitution.

In the case of Brazil, one can argue that the high costs and inefficient methods of production in several of the mature industries (especially those intensive in unskilled labor, the textile industry being an often cited example) resulted from the excessive levels of post-World War II protection. Thus, exposing Brazil to a greater degree of foreign competition by reducing the level of its tariffs may be the single most effective step toward a growth of the country's exports of manufactures. The long-run effect of this measure would be to expand Brazil's exports of manufactures along the line suggested by the principle of comparative advantage.

---

[12] We also experimented with a broader skill index, including workers that earned CR 9,000 or more per month. The results were not materially different from these reported in Table 4.

[13] As regards Brazil's traditional exports of primary commodities, "nature" (i.e., natural resources in conjunction with unskilled labor) has been invoked to explain the country's comparative advantage. Given the unavailability of systematic studies and data on the factor intensity of agricultural commodities, these conclusions are only suggestive. Yet, the limited evidence available points to the combination of soil, climate, and cheap labor as largely accounting for Brazil's exports of such commodities as coffee, cocoa, cotton, tobacco, and sugar. A cogent description of coffee production methods in Brazil is contained in a study undertaken by the FAO.

REFERENCES

W. Baer, *Industrialization and Economic Development in Brazil,* Homewood 1965.

R. E. Baldwin, "Determinants of the Commodity-Structure of U.S. Trade," *Amer. Econ. Rev.*, Mar. 1971, *61*, 126–46.

G. S. Becker, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education,* New York 1964.

J. Bergsman, *Brazil: Industrialization and Trade Policies,* New York 1970.

M. Carnoy, "Rates of Return to Schooling in Latin America," *J. Hum. Resources,* summer 1967, *2,* 359–74.

F. Harbison and C. A. Myers, *Education, Manpower, and Economic Growth,* New York 1964.

G. C. Hufbauer, "Factor Endowments, National Size and Changing Technology: Their Impact on the Commodity Composition of Trade in Manufactured Goods," Univ. Nat. Bur. Econ. Res., Conference on Technology and Competition in International Trade, New York, Oct. 11–12, 1968.

H. G. Johnson, *Comparative Cost and Commercial Policy Theory for a Developing World Economy,* Stockholm 1968.

D. Keesing, "Labor Skills and International Trade: Evaluating Many Trade Flows with a Single Measuring Device," *Rev. Econ. Statist.*, Aug. 1965, *47,* 287–94.

H. B. Lary, *Imports of Manufactures from Less Developed Countries,* New York 1968.

N. H. Leff, "Export Stagnation and Autarchic Development in Brazil, 1947–1962," *Quart. J. Econ.*, May 1967, *81,* 286–301.

———, (1968a) *Economic Policy-Making and Development in Brazil, 1947–1964,* New York 1968.

———, (1968b) *The Brazilian Capital Goods Industry 1929–1964,* Cambridge 1968.

W. Leontief, "Domestic Production and Foreign Trade; The American Capital Position Re-examined," *Amer. Philosophical Soc. Proc.*, Sept. 1953, *97,* 331–49.

T. W. Schultz, "Reflections on Investment in Man," *J. Polit. Econ.*, Oct. 1962, supp., *70,* 1–8.

H. Waehrer, "Inter-industry Skill Differences, Labor Earnings and U.S. Foreign Trade, 1960," unpublished doctoral dissertation, Columbia Univ. 1966.

Instituto Brasileiro de Geografia e Estatístca (IBGE), Recenseamento Geral do Brasil, *Censo Industrial de 1960, Brasil VII,* Rio de Janeiro, GB, Mar. 1967.

International Labour Office, *Yearbook of Labour Statistics 1966,* Geneva 1967.

International Monetary Fund, *Annual Reports on Exchange Restrictions,* various years, Geneva.

United Nations, Economic Commission for Latin America, and the Food and Agricultural Organization; *Coffee in Latin America II. Brazil, State of Sao Paulo, 1,* Mexico 1960.

United Nations, *Yearbook of National Accounts Statistics 1966,* New York 1967.

———, *Commodity Trade Statistics 1966,* New York 1967.

Consumer's Surplus and Producer's Goods

By RICHARD SCHMALENSEE*

We economists generally regard ourselves as scientists, but often we work in the manner of engineers, using a crude approach for an applied problem on the grounds that a more exact procedure is not feasible. Although engineers often use approximations, the careful engineer tries to be aware of and to allow for the properties of the approximations he employs. This essay is an engineering paper in the sense that it examines some characteristics of a crude welfare measure often encountered in applied economics.

In many benefit-cost studies, it is necessary to value a change in price; some measure of the socially equivalent income change is required. In the case of consumer goods, the measure most commonly employed is based on changes in the Marshallian surplus, which is given by changes in the areas under individuals' demand curves for the good in question. It is well known that this is not an exact measure of any carefully defined equivalent income change for individual consumers.[1] The approximation is carried one step further: the change in the area under the market demand curve for the good in question minus the change in the social cost of producing the good is used to measure the change in community welfare. This ignores shifts in the distribution of real income by implicitly assuming that the social marginal utility of income is the same for all individuals.[2]

Seriously flawed though it is, the change in the area under a market demand curve is often the only available measure of community welfare change. Economists, acting in the best engineering tradition, often employ such areas as crude welfare measures in applied work.[3] The purpose of this note is to refine the use of this tool.

Specifically, we shall be concerned with the valuation of changes in the prices of intermediate or producer's goods.[4] We shall assume that the area under the market demand curve for a *consumer* good can be used to construct an acceptable welfare measure. This assumption does *not* allow us to treat firms' demand curves for intermediate goods on the same basis as households' demand curves for consumer goods. A reduction in the price of a producer's good adds to community welfare only to the extent that producer profits and consumer's surplus enjoyed by consumers of the final product increase.

In the next section we shall consider the case of a fall in the price of an intermediate good sold only to a perfectly competitive consumer good industry. Section II analyzes the case where the intermediate good is sold only to a monopolist. In both situations, we assume for simplicity that production takes place under constant returns to scale.[5] Our general conclusion is that demand curves for intermediate goods must be used with extreme care in deriving measures of welfare change.

Before proceeding with the analysis, two

---

* Assistant professor of economics, University of California, San Diego. I am indebted for comments and suggestions to Paul MacAvoy, Wolfhard Ramm, and two anonymous referees. Remaining errors are, of course, my own responsibility.

[1] See Paul Samuelson, pp. 195–202 for a critical comparison of areas under individuals' demand curves with more carefully specified measures of consumer's surplus. These more precise measures are rarely observable, of course. John Hicks (ch. 8–10) and Herbert Mohring, pp. 256–61, have argued that there are situations of practical importance in which the change in the area under an individual consumer's demand curve is a good measure of changes in his consumer's surplus.

[2] See J. de V. Graff, pp. 112–15, on this point.

[3] See Roland McKean (ch. 10) for a general discussion. This measure is used by Jack Hirshleifer, James C. DeHaven, and Jerome W. Milliman (ch. 10), and Paul MacAvoy, among others.

[4] This problem is mentioned by Alan Prest and Ralph Turvey, p. 691. Their conclusion that the demand curves for such goods may be employed on the same footing as demand curves for consumer goods if ". . . the optimum welfare conditions are met all along the line" is not quite correct, as we shall show.

[5] Also, all factor prices and cost functions are assumed to exclude rents. See Ezra Mishan on this point.

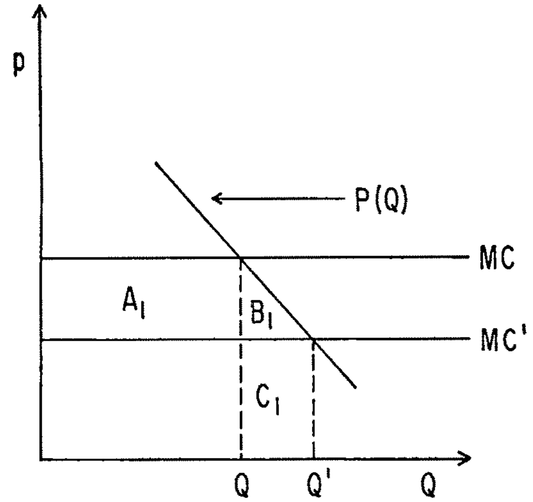


FIGURE 1. TRUE WELFARE GAIN: COMPETITION

properties of efficient production under constant returns to scale must be mentioned. The first is the well-known fact that marginal cost, which equals average cost at all levels of output, depends only on the prices of the factors of production. The second proposition is less well known.[6] Let Q be the output of a good produced under constant returns to scale according to the production function

$$(1)\qquad Q = F(X_1, \ldots, X_n),$$

where X_i is the volume of services of the ith factor. Assume that one unit of X_i costs r_i dollars. Denoting the marginal cost of production by MC, the needed proposition is[7]

$$(2)\qquad \frac{\partial MC}{\partial r_i} = \frac{X_i}{Q}$$

I. The Competitive Case

We shall consider a perfectly competitive industry whose linear homogeneous production function is given by equation (1). To simplify the notation, we shall write X_1 as X and r_1 as r. We are concerned with the welfare gain realized by the community when r falls by a small amount dr to $r' = r - dr$, with r_j constant for $j = 2, \ldots, n$. Let $P(Q)$ be the demand curve for the industry's output, a consumer good.

The fall in r will lower marginal cost from MC to MC' in Figure 1. Using the assumption that the area under the demand curve for the consumer good is a valid welfare measure, the true welfare gain in the competitive case, $TWGC$, is equal to the sum of the areas A_1 and B_1.[8] Utilizing the fact that marginal cost is determined only by factor prices, $TWGC$ may be written as

$$TWGC =$$

$$(3)\qquad Q\frac{\partial MC}{\partial r}dr - \frac{1}{2}\frac{dQ}{dP}\left[\frac{\partial MC}{\partial r}\right]^2 dr^2$$

Now consider Figure 2, which depicts the market for the factor X. A fall in the price of this factor from r to r' increases the demand from X to X'. An analyst who ignored the fact that he was dealing with a producer's good would estimate the welfare gain as the sum of areas A_2 and B_2. We call this estimate $EWGC$, the estimated welfare gain in the competitive case; it may be written as

$$(4)\qquad EWGC = Xdr - \frac{1}{2}\frac{dX}{dr}dr^2$$

Comparing equations (3) and (4) and em-

---

[6] Indeed, I have been unable to find this proposition stated (or proven) anywhere.

[7] The proof of this proposition is quite simple. Under constant returns to scale, average cost equals marginal cost, and we may write

$$MC = (1/Q)\sum_{j=1}^{n} r_j X_j$$

for any output level Q. Assume that Q is fixed and differentiate this equation partially with respect to r_i. This yields

$$\frac{\partial MC}{\partial r_i} = (1/Q)\left[\sum_{j=1}^{n} r_j\frac{\partial X_j}{\partial r_i} + X_i\right]$$

Cost minimizing production requires that $r_j = MC \cdot F_j$ for all j, where F_j is the marginal product of the jth factor. Substituting for r_j, we have

$$\frac{\partial MC}{\partial r_i} = (1/Q)\left[MC\sum_{j=1}^{n} F_j\frac{\partial X_j}{\partial r_i} + X_i\right]$$

Since we are evaluating MC at a fixed output level, the summation term vanishes, and equation (2) in the text is verified.

[8] This is equal to the increase in the area under the demand curve $(B_1 + C_1)$ minus the change in production cost $(C_1 - A_1)$.

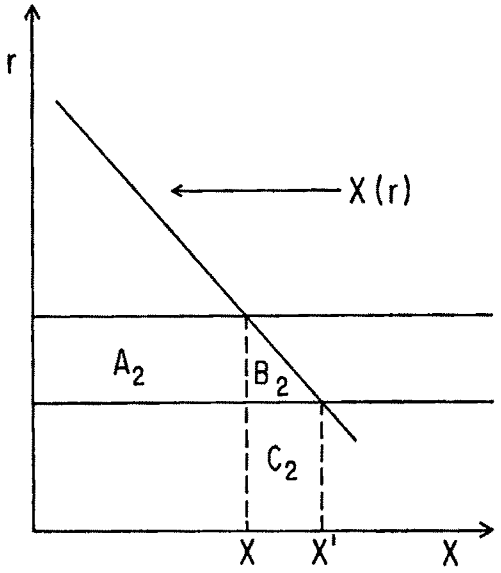


FIGURE 2. ESTIMATED WELFARE GAIN: COMPETITION AND MONOPOLY

ploying equation (2), we have the result that $TWGC = EWGC$ if terms in dr^2 are neglected. For very small changes, it does not matter if the good considered is a producer's good or a consumer's good, as long as it is purchased only by consumers and by perfectly competitive consumer good industries.

To examine the terms in equations (3) and (4) that involve dr^2, we must explicitly consider the demand function for X:

$$X = QG(r, r_2, \ldots, r_n) \tag{5}$$

The form of equation (5) is dictated by the assumption of constant returns to scale: if factor prices are constant, doubling output must double the demand for each input. Since $\partial X/\partial Q = G = X/Q$, we can write

$$\frac{dX}{dr} = \frac{\partial G}{\partial r} Q + \frac{X}{Q} \frac{dQ}{dMC} \frac{\partial MC}{\partial r} \tag{6}$$

Under competition, price equals marginal cost, and $dQ/dMC = dQ/dP$. Substituting (6) into (4), employing (2), and subtracting, we obtain

$$EWGC - TWGC = -\frac{Q}{2} \frac{\partial G}{\partial r} dr^2. \tag{7}$$

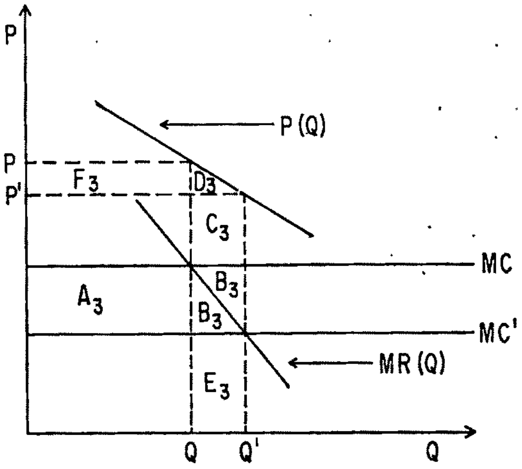


FIGURE 3. TRUE WELFARE GAIN: MONOPOLY

Equation (7) indicates that if the ratio of X to Q, which we called G, is technically fixed, $EWGC = TWGC$ even when terms involving dr^2 are considered. Otherwise the approximation is good only to the first order. Since $\partial G/\partial r$ will be negative, there is a tendency for $EWGC$ to over-estimate $TWGC$. Hence the greater the substitution possibilities for an intermediate good, the greater the error made in using changes in the area under the demand curve for that good as a measure of changes in consumer's surplus. The error is proportional to dr^2 for small price changes, and some defense can be made for ignoring it in applied work.

II. The Monopoly Case

We now consider the impact of a fall in r on community welfare when the intermediate good is used by a firm which is a monopolist in the final market. The production function of the firm is given by equation (1), and all other factor prices are still assumed fixed. The true welfare change is depicted by Figure 3. As a consequence of a fall in r, marginal cost drops from MC to MC'. Marginal revenue, MR, which is always equal to MC in the monopoly case, falls the same amount. Price drops from P to P'. The true gain in welfare, $TWGM$, is given by the sum of the areas A_3, C_3, and D_3, plus twice the area B_3.[9]

---

[9] This is obtained as the increase in the area under the demand curve ($D_3 + C_3 + 2B_3 + E_3$) minus the attendant change in cost ($E_3 - A_3$). Equivalently, it may be viewed

Evaluating these areas in turn, summing, and applying equation (2), we obtain

$$TWGM = \left[1 - \left(\frac{P-MC}{Q}\right)\frac{dQ}{dMR}\right]Xdr \tag{8}$$
$$- \left[\frac{X}{Q}\right]^2 \frac{dQ}{dMR}\left[1 - \left(\frac{1}{2}\right)\frac{dQ}{dMR}\,\frac{dP}{dQ}\right]dr^2$$

Figure 2 applies in the case of monopoly as it did under competitive assumptions.[10] The estimate of welfare gain that would be obtained by considering the market for the intermediate good as if it were the market for a consumer good is given by the sum of A_2 and B_2. Call this measure $EWGM$. It is formally identical to $EWGC$:

$$EWGM = Xdr - \frac{1}{2}\,\frac{dX}{dr}\,dr^2 \tag{9}$$

Let $TWGM'$ be the first-order approximation to $TWGM$ obtained by deleting the term involving dr^2. Define $EWGM'$ similarly. Since marginal revenue falls with output, $EWGM'$ will be less than $TWGM'$. The first-order approximation obtained from the market for the intermediate good understates the true first-order approximation. We shall now consider the extent of the understatement for two simple demand equations.

We first note that for any demand function $P=P(Q)$ we obtain directly

$$\frac{dMR}{dQ} = 2\,\frac{dP}{dQ} + Q\,\frac{d^2P}{dQ^2} \tag{10}$$

Define E as the negative of the price elasticity of demand. For a determinate monopoly equilibrium to exist, E must be greater than one. At the point of monopoly equilibrium, it is well known that

$$\frac{P - MC}{P} = \frac{1}{E} \tag{11}$$

Using equations (10) and (11), we obtain

$$\frac{TWGM'}{EWGM'} = \frac{3 - W}{2 - W}, \tag{12}$$

where

$$W = \frac{EQ^2}{P}\,\frac{d^2P}{dQ^2}$$

If the demand equation for the final (consumer) good is linear, W is zero, and the ratio above is simply 3/2. In this case, to a first approximation, the understatement amounts to one third of the true welfare gain. If the price elasticity of demand is constant, so that $Q=kP^{-E}$, we obtain $W=(E+1)/E$ and

$$\frac{TWGM'}{EWGM'} = \frac{2E - 1}{E - 1} \tag{13}$$

This ratio is greater than one for E greater than one, and it falls as E becomes larger.

Notice that neither of the ratios discussed above falls to one as demand becomes infinitely elastic. The controlling factor is the exercise of monopoly power, not the elasticity of demand for the final good. To see this, consider a situation where pricing is between the extremes of monopoly and competition. Specifically, assume

$$\frac{P - MC}{P} = m\left(\frac{P^* - MC}{P^*}\right), \tag{14}$$

where P^* is the monopoly price. The constant m is assumed to be between zero and one; it may be thought of as the extent to which potential monopoly power is realized. Under this assumption, we must replace dQ/dMC in $TWGM'$ by $dQ/dMC=(dQ/dP)(dP/dMC)$; see equation (8). When the demand for the consumer good is linear, equation (14) implies

$$\frac{TWGM'}{EWGM'} = \frac{2m + 1}{m + 1} \tag{15}$$

This ratio equals one when $m=0$, and it rises to 3/2 as m rises to one. Similarly, if the price elasticity of demand for the consumer good is constant, equation (14) implies

---

as the change in Marshallian consumer's surplus (F_3+D_3) plus the change in monopoly profits ($C_3-F_3+A_3+2B_3+C_3$).

[10] It should be clear that the factor demand curve, $X(r)$, will be different under monopoly than under competitive assumptions. The monopolist's demand curve will lie to the left of that of a competitive industry.

$$\frac{TWGM'}{EWGM'} = \frac{E(m+1) - m}{E - m} \tag{16}$$

This ratio is unity when $m=0$, and it becomes equation (13) when $m=1$.

We now consider the terms in *TWGM* and *EWGM* that involve dr^2. Under monopoly, equation (6) becomes

$$\frac{dX}{dr} = \frac{\partial G}{\partial r} Q + \frac{X^2}{Q} \frac{dQ}{dMR}, \tag{17}$$

employing equation (2). Substituting (17) into (9), we may rewrite the welfare gain estimated from the market for the intermediate good as

$$EWGM = Xdr - \frac{1}{2}\left(Q \frac{\partial G}{\partial r} + \frac{X^2}{Q} \frac{dQ}{dMR}\right) dr^2 \tag{18}$$

Notice that, as in the competitive case, the larger (in absolute value) $\partial G/\partial r$ the larger will be the estimated welfare gain relative to the true gain. In the competitive case, this effect of substitution possibilities imparted an error of second order to the estimated welfare gain, *EWGC*. Here the substitution term raises *EWGM* and reduces the underestimate of welfare gains.

To obtain further results, we shall assume that $\partial G/\partial r=0$. That is, we assume that the ratio of X to Q is strictly fixed by technology. Define $TWGM''$ as the term in *TWGM* involving dr^2, and define $EWGM''$ similarly. Then, under the fixed proportions assumption it can be shown that

$$\frac{TWGM''}{EWGM''} = \frac{3 - 2W}{2 - W}, \tag{19}$$

where W is defined by equation (12). In the linear case, $W=0$, and equation (19) shows that the correction factor given above for this case is unchanged when terms involving dr^2 are considered. It is quite easy to show that equation (15) is also correct when those terms are considered.

In the general case, the ratio of *TWGM* to *EWGM* will lie between the ratios shown in (12) and (19). When the price elasticity of demand for the consumer good is constant $W=E(E+1)/E$, we have

$$\frac{TWGM''}{EWGM''} = \frac{E - 2}{E - 1}, \tag{20}$$

which is less than the ratio shown in equatio (13) for all positive E. Thus, use of (13) o (16) as correction factors will exaggerate th extent to which *EWGM* underestimate *TWGM*. For small changes in r, the error involved in using these factors will be proportional to dr^2, and it would not be unreasonable to ignore it in some applied work.

III. Concluding Comments

This note was not written with the intention of defending the use of areas under market demand curves as welfare measures. Indeed, our chief aim has been to point ou limitations of these areas when producer' goods are considered.

Neither should the result that *EWGM* i less than *TWGM* to a first-order approximation be construed as a defense of monopoly The community will generally be better off i the final good is produced by a competitiv industry, and it will generally realize a large welfare gain from a fall in r if the intermediate good is sold to a competitive industry For instance, if the demand for the fina product is linear and if the ratio of X to (which we have called G) is fixed by technology, it can be shown that $TWGC=(4/3$ *TWGM*. In this simple case the community i better off by a third under competition.

The main result of Section I is that if good or service is sold only to consumers an to perfectly competitive consumer good industries, the good or service can, to a firs approximation, be treated as a consume good for the purposes of consumer's surplu analysis. If areas under demand curves ar valid welfare measures for consumer goods they are valid to a first approximation i this case. This measure will also be correc for price changes where dr^2 cannot be neglected if, and only if, the production function of the consumer good industry is of th fixed-coefficient variety.

In Section II we pointed out that if th

intermediate good is sold to a firm with monopoly power in the consumer good market, it is not valid to treat the demand curve for the intermediate good the same as one would handle the demand curve for a consumer good. Doing so will lead, to a first approximation, to an underestimate of welfare gains. This result generalizes: if monopoly power is exercised in any of the markets that intervene between the market considered and consumers, straightforward use of consumer's surplus analysis in the market for the intermediate good will lead to an understatement of welfare gains.

Section II derived approximate correction factors for two simple demand structures, and we now illustrate their use by an example. Suppose the cost of a particular good can be lowered $1 by a government project costing a round billion dollars. Consumers' and producers' aggregate demand curves for the good are estimated, and the changes in the areas beneath them are computed. It is found that the benefit to consumers would be $600 million, while the area under the producers' demand curve would rise by $300 million. (Both figures are net of increases in production cost.) Suppose that final demand curves are linear. Equation (19) then says that the project should be adopted if *m*, as defined by equation (19), is greater than or equal to one half. At least half of the potential monopoly power in the final product market must be exerted.

One final warning should be stated. As emphasized at the start of this essay, an implicit assumption of aggregative consumer's surplus analysis is that the social marginal utility of income is equal for all individuals. We have retained this assumption here, including the monopolist in the community. But it should be noted that some of the welfare gains that are underestimated in the monopoly case are in fact increases in monopoly profits. We have been assuming that society values such increases equally with other welfare gains, but this is a questionable assumption.

REFERENCES

J. de V. Graff, *Theoretical Welfare Economics,* Cambridge 1967.

J. R. Hicks, *A Revision of Demand Theory,* Oxford 1956.

J. Hirshleifer, J. C. DeHaven, and J. W. Milliman, *Water Supply,* Chicago 1960.

P. W. MacAvoy, *Economic Strategy for Developing Nuclear Breeder Reactors,* Cambridge 1969.

R. McKean, *Efficiency in Government Through Systems Analysis,* New York 1959.

E. J. Mishan, "What is Producer's Surplus?" *Amer. Econ. Rev.,* Dec. 1968, *58,* 1269–79.

H. Mohring, "Urban Highway Investments," in R. Dorfman, ed., *Measuring Benefits of Government Investments,* Washington 1965.

A. R. Prest and R. Turvey, "Cost-Benefit Analysis: A Survey," *Econ. J.,* Dec. 1965, *75,* 683–735.

P. A. Samuelson, *Foundations of Economic Analysis,* Cambridge 1963.

Import Controls on Foreign Oil: Tariff or Quota?

By GEORGE A. HAY*

During the recent deliberation of the Cabinet Task Force on Oil Import Controls, considerable attention was devoted to the question whether, regardless of the level of imports which should be allowed, the control should take the form of a quota or a tariff. Occasionally reference was made to a well-known equivalence theorem[1] which holds that when there is perfect competition in the product market a tariff will generate a volume of imports which, if alternatively set as a quota, produces an identical discrepancy between foreign and domestic prices. Therefore, other than the transfer of revenues from quota holders to the treasury, there should be no difference between the two types of control.

Nevertheless, some support for a quota persisted on the grounds that "A tariff can't do anything to help the consumer. It will have the same effect in limiting imports as a quota, but all the money goes into the treasury. With quotas, on the other hand, there's some chance of benefits trickling down to consumers in the form of lower prices."[2] Since the market conditions and import regulations for oil differ somewhat from the textbook examples, it is worthwhile to consider whether, in this particular case, there is some reason for consumers to prefer quotas to an "equivalent" tariff. It will turn out that due to the peculiar nature of oil quotas which are set as a fraction of refinery output, such a proposition can be shown to be true under a very restrictive set of assumptions.

To set up the problem, it is useful to make some preliminary observations and simplifying assumptions. First, it is approximately true that both the "offer" curv of domestic producers (at the wellhead) an the supply curve for foreign oil delivered t the East Coast of the United States ar horizontal for broad segments.[3] Second, w will assume that all refiners can be characterized as being located along a straigh line from the single producing area, A, t the East Coast port of entry for foreig crude oil, B. Refinery capacity need not b evenly distributed along the line; indeed i is simplest to assume that the concentratio of refining capacity between any two point is related to the amount of consumer demand in the surrounding region. Furthermore, we will assume that demand for crud oil by the refineries in any neighborhoo (and therefore total demand by all refineries) is completely inelastic with respec to price. Finally per barrel transportatio costs either from A, eastward, or from B westward, are assumed proportional to distance traveled.

The government wishes to limit import to a fixed fraction g of total consumption It can do this directly by means of a quota or it can find a tariff which generates tha level of imports. If the latter course i chosen, the government must locate th point W on AB such that a fraction g of th total demand lies to the right of W. (Se Figure 1.) The unit tariff T is then chose

---

* Assistant professor of economics at Yale University. I am grateful to Morris Adelman and Edwin Truman for helpful comments.

[1] See Jagdish Bhagwati.

[2] Former Secretary of Interior Walter Hickel, quoted in *The Wall Street Journal.*

[3] The former is due primarily to the practice o "prorationing" whereby state regulatory commission in the large producing areas control output to kee price constant in the face of month-to-month shifts i demand. Although in the short run a lower price migh cause some marginal producers to shut down (and i the long run it might even reduce the amount of investment in domestic production) this could be completely offset by increased output from the large efficient wells which because of the state-imposed output constraint operate substantially below the poin at which marginal cost equals price. However, thi argument is meant to hold within a fairly narrow range For a sizeable reduction in price, the output constrain would no longer be binding, and the total quantit offered would decline.

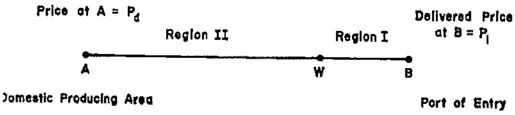


FIGURE 1

so that at W:

$$(1)\quad P_d + s(W - A) = P_i + T + s(B - W)$$

where:

P_d = domestic price at A
P_i = foreign price delivered at B
s = transport cost per barrel-mile

Therefore:

$$(2)\quad T = P_d - P_i + s[(W - A) - (B - W)]$$

Refiners in Region I (to the right of W) will import all their requirements at a cost of $P_i + T$ plus the inland transportation costs from B. Refiners in Region II (to the left of W) will pay P_d plus the transportation costs from A.

Under the quota system as it is currently administered,[4] each refiner is permitted to import a fraction g of his production. However, since "tickets" are relatively worthless to refiners near A because of the inland transportation costs, these refiners are permitted to "sell"[5] their tickets to refiners closer to B. It is clear that under a well-functioning arbitrage system, the end result should again be that all the imported foreign oil still winds up in Region I. A fraction g of the oil would be available under Region I's quota at a price P_i, but the rights to the remaining imports would have to be purchased from Region II refiners at some markup P_q over P_i.

We now wish to compare the per barrel cost to refiners in Region I under the alternative control systems.

Under a tariff:

$$(3)\qquad C_t = P_i + T + ITP$$

Inland transportation costs (ITP) are a function of the dispersion of demand within Region I but need not be specified in detail because they will be the same under a quota system so long as the arbitrage system described above works smoothly.

Under a quota:

$$(4)\quad C_q = gP_i + (1 - g)(P_i + P_q) + ITP$$

By rewriting (3) as:

$$(3')\qquad C_t = g(P_i + T) + (1 - g)(P_i + T) + ITP,$$

we can see more clearly the relative costs of the alternative controls. On the one hand, Region I refiners under a quota regime save an average of gT per barrel by being allowed to import a fraction g of total requirements with no tariff. On the other hand, Region I refiners lose $(1-g)(P_q-T)$ which is the premium paid for Region II's import tickets on a fraction $(1-g)$ of total requirements over the amount that Region I refiners would pay under a tariff.

What is P_q likely to be? If the market for tickets is perfectly competitive (although the total number of tickets is fixed) it is easy to show that $P_q = T$ and therefore that $(1-g)(P_q-T) = 0$.

In a perfectly competitive market, the price of a ticket will have to be low enough so that none of Region II's tickets go unsold, yet high enough so that Region II refiners will not prefer to utilize the import tickets for their own production rather than sell to Region I. This will occur if a refiner at W—the border between Region I and Region II—is indifferent between using a ticket himself and selling it.

This in turn requires that:

$$(5)\quad P_i + s(B - W) = P_d + s(W - A) - P_q$$

or

$$(5')\qquad P_q = P_d - P_i + s[(W - A) - (B - W)]$$

$$(6)\qquad \because P_q = T, \quad \text{and} \quad C_t - C_q = gT$$

---

[4] With minor modifications discussed below.

[5] "Imported, crude and unfinished oils must be processed in the licensee's refinery or petrochemical plant or exchanged for domestic crude or unfinished oils which are so processed by the licensee. Sales are thus prohibited, but oil companies have developed a system of exchanges utilizing variable exchange ratios which permits licensees to realize an effective financial benefit rom the value of the import license" Task Force Report, p. 15.

Therefore, under the assumption of a perfect market for import tickets, refiners in Region I pay out on average gT per barrel less than they would under a tariff system. Similarly, refiners in Region II benefit under a quota system by an amount equal to $(1-g)T$ times Region I's requirements. The net effect, therefore, of replacing a quota with a tariff is to increase the federal revenues by an amount equal to T times Region I's consumption,[6] with refiners in each region "contributing" an amount directly proportional to the region's share of total demand.[7]

Furthermore, since it is consumer not refiner welfare in which we are interested, it is essential to note that the saving to refiners under a quota system tends to reduce *marginal*, not merely average, costs since each refiner's quota allocation is proportional to his expected output. Thus, *at the margin*, the refiners' unit costs are increased by gT if the quota system is replaced by a tariff, and we would expect oil prices to increase. Therefore, under the conditions of this example, the increased federal revenues are not in fact a transfer from refiners but come ultimately from oil consumers in the form of higher prices.

This is in contrast to the usual textbook situation in which quota allocations are fixed independently of production. In that case, the *marginal* cost to refiners under a quota would be no lower than under a tariff, and there is no reason to expect a pass through of the savings which refiners collect on the barrels imported under the quota.[8]

I. Qualifications

The above result depends on the existenc
of a competitive market in import ticket
To the extent that tickets are not price
competitively,[9] Region II can push up th
costs for Region I refiners. In particular,
the supply of tickets in Region II could b
monopolistically controlled, it would b
possible to discriminate perfectly in the sal
of tickets to Region I charging a price t
each refiner which would make him in
different between buying tickets and usin
domestically produced oil. For a refine
located at B, for example, the ticket pric
would be set at:

$$P_q = P_d + s(B - A) - P_i \tag{7}$$

Depending on the dispersion of Region I'
refineries along the segment BW, it is en
tirely possible that the total cost to Region
refiners in this situation would be sub
stantially greater than it would be under
tariff. In any case the relative savings to th
two regions with a quota system are sub
stantially changed, with Region II comin
out ahead.

In addition, we have assumed above tha
changes at the margin in unit raw material
cost to refiners will be passed through t
consumers in lower prices. Obviously, i
there is not perfect competition in the re
fining industry, some of the lower costs o
oil under a quota would be absorbed by th
refiners, and the preference for a quota i
weakened. This is reinforced by the fact tha
the marginal quota allotment to a refine
declines somewhat at higher outputs.[10]

Furthermore, we have ignored the obviou
fact that under a tariff system, the mone
collected by the treasury is not a complet
loss to consumers since oil customers ar
also taxpayers, and with government ex
penditures held constant, the tax bill woul
be lowered. However, unless the use of oil i

---

[6] In the United States in 1969, this would have been something in the order of \$2 billion, since g is .122, and estimates of the "equivalent" tariff went as high as \$1.75 per barrel. In mid-1970, a short-run tanker shortage caused by events in the Middle East rendered import tickets temporarily worthless.

[7] Using this analysis, it is easy to see that an alternative quota system under which all the tickets were allocated directly to Region I would not change the "real" results but would result in a transfer from Region II to Region I of $(1-g)T$ times Region I's requirements as compared with the standard quota. Region I should clearly prefer this type of arrangement to the standard quota.

[8] That is, the lower cost of the "quota barrels" would be appropriated by the refiners as a pure rent.

[9] Or if the arbitrage process does not work efficientl
so that Region I refiners wind up using some domesti
crude oil while some Region II refiners do not sell thei
tickets and actually use the foreign oil.

[10] The Task Force Report p. 22, contains a "guesti
mate" that no more than half the savings are passed o
to consumers.

distributed approximately in proportion to individuals' tax payments (an assumption which is clearly not valid) an across-the-board tax cut to "refund" the tariff revenues would be relatively small consolation to oil users.

Finally, even if all the assumptions most favorable to a quota system are fulfilled, there may still remain strong reasons to favor replacing the quota with a tariff. First, conditions in domestic and world oil markets plus developments in construction of oil tankers make it likely that the tendency will be for the prices of domestic and (delivered) foreign oil to continue to drift further apart. If the tariff is not continuously revised upward to offset this development, or is revised only with a lag, consumers derive some advantage. Looked at another way, the incentive for domestic producers to initiate price increases might be somewhat dampened since each price hike would require a simultaneous request for an increase in the tariff. Second, the administration of the quota system leaves sufficient room for "discretionary" allotments not necessarily based on current refinery output, that incentive for using political influence or bribery to obtain favorable treatment poses a constant threat to the integrity of the system.[11] The tariff system appears to offer a considerable improvement in this regard.

II. Postscript

A broader and more fundamental issue has not been considered in this paper, although I have dealt with it at length elsewhere. Specifically, what is the rationale for imposing any form of import control on crude petroleum and petroleum products, especially at the restrictive levels currently involved? Both my own estimate and that of the Task Force indicate that the cost to petroleum users of the import control program is of the order of $5 billion in 1968. It is not obvious that the contribution to national security of import controls is sufficient to justify such a cost.

REFERENCES

J. Bhagwati, "On the Equivalence of Tariffs and Quotas," in R. E. Baldwin et al., eds., *Trade, Growth and the Balance of Payments: Essays in Honor of Gottfried Haberler,* Chicago, 1965.

K. Dam, "Implementation of Import Quotas: The Case of Oil," *J. Law Econ.,* Apr. 1971, *14,* 1–60.

G. Hay, *The New England Council Submission to the Cabinet Task Force on Oil Import Controls,* July 1969.

The Cabinet Task Force on Oil Import Controls, *The Oil Import Question,* 1970.

Wall Street Journal, Dec. 5, 1969.

---

[11] There is no implication that any dishonest or illegal activity actually takes place, merely that the incentives to cheat are substantial. See the recent paper by Kenneth Dam.

Profits of Defense Contractors

By George J. Stigler and Claire Friedland*

The profitability of the large defense contractors has been a popular topic in the public discourses on the military establishment. Excessive reliance has been placed on striking examples of hugely successful or fatally unsuccessful enterprises. We have been led to make a more systematic examination of the profitability record because one of us served on a Blue Ribbon Defense Panel, which recently submitted a report on the organization of the Department of Defense.

First a report on this subject by the Logistics Management Institute, a research arm of the Department of Defense, may be summarized. The Institute requests of some forty major defense contractors a separation of assets and net income between "defense" and "commercial" business. The main result is presented in Table I.[1] In the late 1950's, the defense business was more profitable than either the commercial business of the same companies or the Federal Trade Commission-Securities and Exchange Commission (*FTC-SEC*) universe of 3500 companies. After 1961, this pattern was reversed.

Table 1—Rates of Return on Total Capital Invested, Defense Contractors, and *FTC-SEC* Quarterly Reports

	Rates of Return (After Taxes)		
	Defense Contractors		
Year	Defense Business	Commercial Business	FTC-SEC
1958	10.1%	7.0%	7.1%
1959	9.5	6.8	9.3
1960	8.7	4.8	7.8
1961	7.5	4.7	7.4
1962	7.4	9.0	9.3
1963	6.5	8.7	9.8
1964	6.3	10.9	10.8
1965	7.6	11.6	12.6
1966	7.0	10.8	12.4
1967	7.3	7.4	10.1
1968	6.8	8.3	10.2

Source: Defense Industry Profit Review, p. 11.

I. The Investigation of Profit

A study of the profitability of investments in major defense contractors was made for several reasons. Stock market experience avoids (or at least ignores!) the complications of accounting practices, including the difficulties of segregating assets and income within the enterprise. The procedure, in brief, was as follows:

1. The largest defense prime contractors are identified in the annual lists compiled by the Department of Defense. Because the contracts placed vary erratically from year to year, the average of prime contracts over a period of years is used; the volume of defense business would be altered moderately if we could take account of subcontracts let and received.
2. These contracts are compared with total sales of the companies for the same years, to obtain an average percentage of defense business.
3. The market value of an initial investment in the common stock of a company, with reinvestment of all dividends in the stock of the company, is calculated from the Center for Research in Security Prices (*CRSP*) tape.

The data for each company are available on request.

II. Investment Performance

The interpretation of the measure of investment performance in Tables 2 and 3 is as follows: if $1,000 were invested in each company listed upon the New York Stock

---

* University of Chicago.

[1] The criteria for inclusion in 1967, for example, were that the companies each have $25 million or more of defense sales, amounting to 10 percent or more of all its sales.

Table 2—Stock Market Performance of Large Defense Contractors, 1948–61, June 1948 = 1.000 (unweighted averages): Value of Initial Shares Plus Reinvestment of Dividends

Year (June)	All Stocks *NYSE*	1950–57 Large Defense Contractors (54)
1948	1.000	1.000
1949	.820	.991
1950	1.106	1.457
1951	1.464	1.963
1952	1.723	2.545
1953	1.786	2.624
1954	2.134	3.798
1955	3.149	6.198
1956	3.515	7.156
1957	3.741	7.790
1958	3.841	7.769
1959	5.510	11.198
1960	5.488	10.515
1961	6.740	12.573

Table 3—Stock Market Performance of Large Defense Contractors, 1958–68, June 1958 = 1.000 (unweighted averages): Value of Initial Shares Plus Reinvestment of Dividends

Year (June)	All Stocks *NYSE*	Large Defense Contractors	
		1959 Top Contractors (50)	1969 Top Contractors (48)
1958	1.000	1.000	1.000
1959	1.434	1.449	1.600
1960	1.429	1.356	1.648
1961	1.755	1.618	1.945
1962	1.442	1.378	1.581
1963	1.869	1.793	2.121
1964	2.148	2.081	2.392
1965	2.386	2.252	2.683
1966	2.902	2.970	3.978
1967	3.596	3.558	4.830
1968	4.702	3.860	5.254

Exchange (*NYSE*) in June of 1948, with reinvestment of dividends, the average value of the shares of a listed company would be $1,106 in June 1950 and $3,841 in June 1958. For each of the two wars, in Korea and Vietnam, we choose a base period in which the outbreak or escalation of hostilities would not be anticipated. The main findings are:

1. The investments in defense contractors were almost twice as profitable in the 1950's as the investments in all listed stocks. For example, $1,000 invested in 1948–49 in the 54 leading defense contractors of 1950–57 (weighted equally) would have been worth $7,816 in 1957–58 compared with a corresponding figure of $4,166 for all *NYSE* stocks.
2. In the 1960's the investments in defense contractors did approximately as well as in all *NYSE* stocks. For example, with the foresight in 1958 to have chosen the leading defense contractors of 1969, a $1,000 investment would have grown to $5,254 in 1968,[2] compared to $4,702 for all *NYSE* stocks. If the investments were in the leading defense contractors of 1959, the 1968 value would have been $3,860.

Thus there is substantial agreement between our study and that of the Logistics Management Institute.

3. In the 1950's the profitability of investing in a defense contractor was greater, the larger the ratio of his defense to total sales. In the 1960's there was no such relationship. Correlating stock market performance with ratio of defense to total sales, there is a positive relationship in the 1950's ($r^2 = .295$, 52 companies), no relationship in the 1960's ($r^2 = .00008$, with the leading defense contractors of 1959.)[3]

---

[2] Alternatively, one may construct a *weighted* average return, using the total value of the common stock outstanding of each company at the end of 1963 for the 1958–68 analysis: this rises from $1,000 in 1958 to $3,629 in 1968 (as compared with the unweighted result, $5,254). Unfortunately, there is no comparable total market performance index with which to interpret this weighted average. (A corresponding calculation was performed for the 1948–58 period, but the difference between unweighted and weighted results was not appreciable.)

[3] In this latter period, average 1964–68 market values are compared with 1958–60 market values, and share of defense sales is calculated for 1958–63; N = 50.

III. Answered and Unanswered Questions

The popular view of unusual profits of defense contractors is therefore correct for the 1950's and incorrect for the 1960's—a better score than popular views usually have. The economist is more concerned with several other questions:

1. Is the defense area riskier? Our tentative answer is in the affirmative: there is a positive correlation between the instability over time of the sales of a company and its share of defense business ($r^2=.337$, 1950–57, 68 companies; $r^2=.245$, 1958–63, 75 companies).
2. Is there substantial monopoly or blocked entry into the defense contracting business?
3. How is the profitability of individual defense contractors related to their performance on contracts?

Until these difficult latter questions are answered more fully than our present knowledge allows, we shall be ill-equipped to judge the efficiency of present systems of procurement.

REFERENCE

Logistics Management Institute, *Defense Industry Profit Review,* 1968 Profit Data, Mar. 70.

On the Empirical Relevancy of the Ranis-Fei Model of Economic Development: Comment

By HAKCHUNG J. CHOO*

> "The words figure and fictitions both derive from the same Latin root, fingere. Beware!"
>
> M. J. Moroney

In recent years, Gustav Ranis and John C. H. Fei (1961, 1963) have advanced a theory of economic development based on the celebrated W. Arthur Lewis model of development with unlimited supplies of labor. Their model of development emphasizes the role of Lewis' neglected subsistence sector (i.e., the agricultural sector) focusing on the Rostovian stage of take off to sustained growth and the impacts of technology on development during this period. Utilizing such familiar, yet controversial concepts as disguised unemployment, subsistence wage, terms-of-trade effects, balanced growth, and others, they have concluded that an underdeveloped country can shift successfully its center of gravity from the agricultural sector to the industrial sector by allocating its investible resources to maintain the balanced-growth path, by maintaining the subsistence wage level and by adopting labor-intensive technology until the "turning point" is reached. Furthermore, Fei and Ranis assert that the adoption of the only successful criterion for resource allocation demonstrates a capital shallowing tendency in the earlier years of development and then a capital deepening trend in the later stage of development, which is not in conflict with the two major social objectives of maximum output and maximum employment. Later, in their monograph (1964) relaxing the initial assumption of a closed economy, they contend that no major modification is necessary as long as the foreign sector is a "facilitator" of growth, but not the "main engine."

To support the model empirically, Fei and Ranis relied exclusively on a comparative study of historical Japan, 1888–1930, and India, 1949–60. According to their findings, successful Japan has demonstrated a pronounced capital shallowing phase and adoption of labor-intensive technology until 1917[1] and a capital deepening phase and adoption of capital-intensive technology thereafter, in contrast to the case of India. If the validity of the Ranis-Fei model is accepted it will continue to have, as it has had, tremendous theoretical and policy implications on development efforts by underdeveloped countries and on the continuing controversies over investment criteria and balanced versus unbalanced growth.

Succeeding the comments by Harry T. Oshima and Edwin P. Reubens, the purpose of this paper is to reexamine and cast doubt on the empirical relevancy of the Ranis-Fei model by scrutinizing the successful Japanese experience in the light of the most recent estimates of capital stock and prices by Kazushi Ohkawa, M. Shinora, and M. Umemura (1966, 1967). The major emphasis will lie in verifying the comparability of basic data for the two countries and in examining whether Japan actually experienced the alleged capital shallowing phase. Throughout I shall follow the statistical and analytical techniques advanced by Fei and Ranis.

---

* Associate professor of economics and statistics and codirector of bureau of business and economic research at Nichols College. I am grateful to the late Eugene Melder of Clark University and Sang C. Suh of the World Bank, advisors of my M.A. thesis, on which this comment is based. I also wish to thank Howard W. Nicholson of Clark University for his suggestion and encouragement to prepare this manuscript; Henry J. Bruton of Williams College; E. E. Hagen of M.I.T.; my colleague Ernest D. Phelps of Nichols College; Edwin P. Reubens of the City University of New York; Morris Singer of the University of Connecticut; and an unknown referee for their valuable comments on the earlier draft of this manuscript.

[1] In his earlier article in this *Review*, Ranis concluded that such a turning point was empirically identified somewhere in 1895. There is no clarification by Fei and Ranis as to why it is identified in 1917 in their later study and how they are related to each other. But I suspect that this difference stems from the fact that his earlier study concentrated on a segment of the Japanese industrial sector, namely textile industry, while in their later study, they dealt with the nonagricultural sector as a whole.

I shall also attempt to verify empirically such key theoretical assumptions and arguments in the Fei-Ranis model as the rigidity of subsistence wage, powerful "terms-of-trade effects" and the underemphasized role of international trade in Japanese development.

I. Fei and Ranis' Data and Method

Fei and Ranis present three tests in evaluating the development efforts of an underdeveloped country to shift its center of gravity toward industrialization. The first test is the Leibenstein critical minimum effort thesis that the industrial sector should absorb more than the net population growth rate either by a sufficient rate of capital formation and/or by promoting labor-intensive technology. The second test is an analysis of the capital-labor ratio over time. A successful transition, according to Fei and Ranis, is supposed to show declining ratios in the earlier years and then increasing ratios after the turning point.

Their third test is a somewhat ingenious contribution. The magnitude or rate of labor absorption by the industrial sector is decomposed into two parts: one due to capital accumulation by the Hicksian assumption of neutral technology (i.e., their "radical effect") and the other due to the bias of labor- or capital-intensive technology being applied by relaxing the Hicksian neutrality assumption (i.e., their horizontal effect). Symbolically, the labor absorption equation reads:

$$\eta_L = \eta_K + \frac{B_L + J}{\epsilon_{LL}} = \eta_r + \eta_h$$

where

η_L = the rate of labor absorption by the industrial sector;
η_K = the rate of capital accumulation;
B_L = the degree of labor-using bias of innovation;
J = the intensity of innovation;
ϵ_{LL} = the elasticity of the marginal physical productivity of labor curve;
η_r = radical effect; and
η_h = horizontal effect.

To provide an empirical context to these three tests, the statistical method used by Fei and Ranis is essentially quadratic regression and correlation analysis of capital stock to "gainfully employed labor" over time. This method is used primarily for the reason that it provides a best fitting and differentiable curve for the purpose of their study. Therefore, the issue is the accuracy and comparability of the estimates of capital stock and employed labor for India and Japan.[2]

II. Qualitative Aspects of the Ranis-Fei Analysis and Data

In applying the three tests to a comparative study of Japan and India, my major objection is that Fei and Ranis seem to emphasize the marginal analysis of capital and labor and changes in capital-labor ratios over time, without considering the initial conditions of the two countries and the resulting mix of resources. By neglecting them, Fei and Ranis implicitly assumed conditions in the two countries (Japan in the 1890's and India in the 1950's) to be more or less the same. It is true that W. W. Rostow classified and placed the two countries in these years in the stage of take off, but it is an oversimplification to assume the initial conditions of the two to be more or less the same. As an example, Henry Rosovsky found that Japanese initial conditions were much more favorable than many underdeveloped countries of today with respect to the initial capital endowment and the supply of educated manpower. Furthermore, the geographical size and relative backwardness of historical India as compared to historical Japan would also have some ramifications upon its allocation of investible resources.

---

[2] In addition, there are the questions of comparability of such exogeneous factors as the time periods being compared, the level of technology, geographical and physical sizes of the two economies and thus, the requirements of social overheads, and other initial conditions and factors. These questions center more or less around the weakness of empirical studies in social sciences where a controlled experiment is not possible. However, an analyst should make some mental notes as to whether or not neglecting such exogeneous factors seriously restrict his findings.

Without due consideration of the initial conditions and their implications for resource allocation in a comparative study, the following conclusions would be meaningless if not erroneous. For example, based on Ranis-Fei data in their article in this *Review* (1963), the average capital-labor ratio of Japanese nonagricultural sector in 1888 and of Indian manufacturing and mining sector in 1949 are 4,272.34 (1928–32) constant yen and 2,017.96 (1953) constant rupee, respectively.[3] Taking the relative purchasing powers of the two currencies into consideration, but neglecting intensified capital requirements of modern technology inevitably resulting from the time gap between the base years of the two nations, the Japanese average capital-labor ratio of 1888 is approximately twice as much as that of India in 1949. If this difference in the initial capital endowment per worker is recognized, is it not rational for India to allocate capital intensively to reach the initial Japanese level and to exploit the late comers' advantages?

Turning to Fei and Ranis' data and estimates used, the acceptability of their empirical tests is even more questionable. First of all, the industrial classifications for the two countries differ from each other: the industrial sector of Japan is the nonagricultural sector (consistent with their two-sector model) while the Indian industrial sector only includes manufacturing and mining, leaving out the tertiary sector (inconsistent with their two-sector model).

Secondly, although Fei and Ranis have used a term, "gainfully employed labor," in their analysis without any distinction, the difference in the statistical concepts of "gainfully employed labor" for Japan and India should not be underemphasized, especially when these data are to be applied to such a sensitive analysis as their "decomposition analysis."

For India, it is mainly the census employment figures of the manufacturing and mining sectors, whereas for Japan, it is not the census data, but Ohkawa and others' "gainfully occupied labor" of the nonagricultural sector. The contrast between the census employment and gainfully occupied labor can be readily seen in the definition of the latter by Ohkawa et al.:

> Gainfully occupied population is . . . a concept which is difficult to handle, and not much in vogue among Western economists. Gainfully occupied population does not mean labor force or actually employed population. Rather it refers to the usual status of the individual in question. In other words, persons are queried about their usual occupations, and are listed accordingly. Therefore, if at the time of the census they happen to be unemployed, in fact they will still be listed as gainfully employed. [1957, p. 142]

Thus, Ohkawa et al. did not fail to add a word of caution to statistical analysts in applying this concept and using these figures by saying:

> . . . as a result, the participation rate of people who work compared to total population is *extremely high* by Western standards . . . in practice, in dealing with gainfully occupied figures their hetrogeneous compositions should always be remembered. [1957, p. 142, italics added.]

It seems evident that this warning was entirely overlooked by Fei and Ranis.

It is, therefore, obvious from the quotations that Japanese nonagricultural labor statistics are significantly inflated by the very definition of gainfully occupied labor compared to Indian manufacturing and mining census labor statistics. Since persons are queried about their usual occupations, it is also likely that the Japanese agricultural gainfully occupied labor in the early years, due to the unemployed and underemployed, might have been inflated more than that in the later years. Conversely, the same bias might have affected the nonagricultural gainfully occupied labor in the later years as industrialization took place and the labor force began to migrate into urban centers seeking better employment opportunities. In short, high portions of the unemployed and underemployed would have answered "agriculture" in the early years and "nonagriculture" in the later years when asked about their usual status of employment regardless of their current employment status. Consequently, there is a strong possibility that this concept of gainfully occupied labor provided built-in biases in favor of the Fei-Ranis thesis by lowering the capital-labor ratios and inflating the labor absorption rates by the industrial sector in the later years in contrast to their reversing effects in the earlier years.

---

[3] Fortunately, there was no significant difference in the purchasing powers of the 1928–32 Japanese yen and the 1953 Indian rupee in terms of a constant *U.S.* dollar. If not, one should have regressed and correlated capital stocks of the two nations in a common denominator to homogeneous employment figures.

As for the data on capital stock for the two countries, Fei and Ranis derived this data in the same manner in the sense that they used the so-called bench mark year method in deriving time-series data. For India, net yearly investment in manufacturing and mining in constant prices is added to U. Datta and V. Prahash's estimates of reproducible tangible wealth in manufacturing and mining, allowing for different rates of depreciation to small and large scale industries. Whereas for Japan, Rosovsky's estimates of five-year moving averages of gross investment in nonresidential construction, durable equipment, and public investment are deflated by Ohkawa's nonagricultural price index. Allowing two alternative depreciation rates, 15 and 20 percent of gross investment, the cumulated net investment is added to M. Ito's estimates of the 1930 capital stock of the nonagricultural sector based on the Japanese national wealth census. It is not the method of estimating annual capital stock that is questionable, but the reliability of capital stock estimates of the bench mark year.

Ohkawa, Shinohara, and Umemura noted that Ito's pioneering estimates of Japanese capital stock on the basis of national wealth censuses are grossly inadequate to apply them to any meaningful analysis (1966, p. 5). Ito's data are overestimates. Consequently, the rate of annual capital stock increase for Japan never exceeded 2 percent until 1920 while the Indian rate ranged from 1.9 percent to 11.8 percent (1963, p. 290, 294) and the average capital-labor ratio of Japan in the initial year was extremely high as noted earlier.

Of course, the difficulties stemming from collecting and analyzing historical data for their purpose at the time of Fei and Ranis' work are understandable and deserve sympathy in the light of restrictiveness and paucity of available data (although there were some twelve series of capital stock estimates, published and unpublished, in Japan at that time). However, what is desired is a more careful assessment of data, and some qualifying reservations on the reliability and comparability of data for the two nations.

III. Historical Japanese Experience Reexamined

Reubens, in his commentary in this *Review*, attempted to demonstrate that historical Japan did not experience a capital-shallowing period in her early stage of development. He calculated the capital-labor ratios for all and for the industrial sector. This effort was not received by Fei and Ranis as conclusive evidence to refute their empirical test because of the fact that Reubens did so by using his adjusted estimates of capital stock. In the present attempt to reevaluate historical Japanese experience, no effort is made to devise estimates of manufacturing and mining capital stock and labor for Japan. I shall avoid similar controversies over estimating procedures, by substituting recent conclusive capital stock estimates of the nonagricultural sector by Ohkawa, Shinora, and Umemura (1966) for those used by Fei and Ranis.

Since there is no way of separating the employment figures compatible to census employment from gainfully occupied figures, except to apply some educated guesses, no attempt is made to separate them in this reexamination. Therefore, it is still biased in favor of the Ranis-Fei thesis in the sense that Japanese industrial labor is inflated by the definition of gainfully occupied population.

By following exactly the same analytical and statistical procedures advanced by Fei and Ranis, the application of the new capital stock estimated by Ohkawa et al. to the gainfully occupied labor in the nonagricultural sector used by Fei and Ranis yields a quadratic regression line.

$$K = 2{,}819.3057 - .849474L + .1288503L^2, \quad \text{with } R^2 = .9964$$

The resulting capital-labor ratio and decomposition analysis[4] are shown in Table 1.

TABLE 1—JAPANESE CAPITAL-LABOR RATIO AND DECOMPOSITION ANALYSIS 1888–1930

Year	Capital-Labor Ratio			Decomposition Analysis		
	Net Capital Stocks[a] (K) (Mill. Yen)	Gainfully Occupied Labor (L) (Thousand)	K/L (Yen)	$\Delta L/\Delta t$ (η_L) (Thous.)	$\Delta r/\Delta t$ (η_r) (Thous.)	$\Delta h/\Delta t$ (η_h) (Thous.)
1888	1,875	4,993	375.52	—	—	—
	(—)	(—)	(—)	(—)	(—)	(—)
89	1,965	5,188	378.75	195	240	−45
	(4.80)[b]	(3.91)[c]	(.86)	(3.91)[c]	(4.80)[b]	(− .89)[d]
90	2,069	5,385	384.21	197	274	−77
	(5.29)	(3.79)	(1.44)	(3.79)	(5.29)	(−1.50)
91	2,183	5,582	391.07	197	297	−100
	(5.51)	(3.65)	(1.78)	(3.65)	(5.51)	(−1.86)
92	2,292	5,781	396.47	199	278	−79
	(4.99)	(3.65)	(1.38)	(3.65)	(4.99)	(−1.34)
93	2,419	5,980	404.51	199	320	−121
	(5.54)	(3.44)	(2.03)	(3.44)	(5.54)	(−2.10)
94	2,576	6,183	416.62	203	388	−185
	(6.49)	(3.39)	(2.99)	(3.39)	(6.49)	(−3.10)
95	2,750	6,384	430.76	201	417	−216
	(6.75)	(3.25)	(3.39)	(3.25)	(6.75)	(−3.50)
96	2,928	6,588	444.44	204	413	−209
	(6.47)	(3.19)	(3.18)	(3.19)	(6.47)	(−3.28)
97	3,116	6,803	458.03	215	423	−208
	(6.42)	(3.26)	(3.12)	(3.26)	(6.42)	(−3.16)
98	3,304	7,000	472.00	197	410	−213
	(6.03)	(2.89)	(3.05)	(2.89)	(6.03)	(−3.14)
99	3,493	7,216	484.06	216	400	−184
	(5.72)	(3.08)	(2.56)	(3.08)	(5.72)	(−2.64)
1900	3,680	7,437	494.82	221	386	−165
	(5.35)	(3.06)	(2.22)	(3.06)	(5.35)	(−2.29)
01	3,835	7,666	500.26	229	313	−84
	(4.21)	(3.08)	(1.10)	(3.08)	(4.21)	(−1.13)

[a] Net Capital stocks in the nonagricultural sector. Seven-year moving averages in 1934–36 constant prices.

[b] Rate of Capital accumulation.

[c] Rate of Labor Absorption.

[d] Horizontal Effect.

Sources: Ohkawa, Shinohara, and Umemura, (1966, pp. 246–49); and Fei and Ranis (1963, pp. 290–91).

---

[4] Fei and Ranis used the incremental ratios of regression values of capital over time in determining the radical effect, without making a specific reference as to

TABLE 1—(*Continued*)

Year	K	L	K/L	$\Delta L/\Delta t$	$\Delta r/\Delta t$	$\Delta h/\Delta t$
02	4,005	7,883	508.05	217	340	−123
	(4.43)	(2.83)	(1.56)	(2.83)	(4.43)	(−1.60)
03	4,200	8,111	517.81	228	384	−156
	(4.87)	(2.89)	(1.92)	(2.89)	(4.87)	(−1.98)
04	4,408	8,331	529.10	240	401	−161
	(4.95)	(2.71)	(2.18)	(2.71)	(4.95)	(−2.24)
05	4,647	8,561	542.81	210	452	−242
	(5.42)	(2.52)	(2.59)	(2.52)	(5.42)	(−2.90)
06	4,926	8,787	549.22	226	514	−288
	(6.00)	(2.64)	(1.18)	(2.64)	(6.00)	(−3.36)
07	5,252	9,012	582.77	225	582	−357
	(6.62)	(2.56)	(6.11)	(2.56)	(6.62)	(−4.06)
1908	5,572	9,234	603.42	222	549	−327
	(6.09)	(2.46)	(3.54)	(2.46)	(6.09)	(−3.63)
09	5,956	9,458	629.73	224	636	−412
	(6.89)	(2.43)	(4.36)	(2.43)	(6.89)	(−4.46)
10	6,373	9,680	658.36	222	662	−440
	(7.00)	(2.35)	(4.55)	(2.35)	(7.00)	(−4.65)
11	6,801	9,907	686.48	227	650	−423
	(6.72)	(2.35)	(4.27)	(2.35)	(6.72)	(−4.37)
12	7,214	10,134	711.86	227	601	−374
	(6.07)	(2.29)	(3.70)	(2.29)	(6.07)	(−3.78)
13	7,645	10,360	737.93	226	605	−379
	(5.97)	(2.23)	(3.66)	(2.23)	(5.97)	(−3.74)
14	8,083	10,590	763.26	230	594	−364
	(5.73)	(2.22)	(3.43)	(2.22)	(5.73)	(−3.51)
15	8,616	10,811	796.96	221	698	−477
	(6.59)	(2.09)	(4.42)	(2.09)	(6.59)	(−4.50)
16	9,250	11,036	838.16	225	796	−571
	(7.36)	(2.08)	(5.17)	(2.08)	(7.36)	(−5.28)
17	9,964	11,263	884.66	227	852	−625
	(7.72)	(2.06)	(5.55)	(2.06)	(7.72)	(−5.66)
18	10,729	11,486	934.09	223	864	−641
	(7.67)	(1.98)	(5.59)	(1.98)	(7.67)	(−5.69)
19	11,545	11,714	985.57	228	874	−646
	(7.61)	(1.98)	(5.51)	(1.98)	(7.61)	(−5.63)
20	12,352	12,415	998.92	701	819	−118
	(6.99)	(5.98)	(.95)	(5.98)	(6.99)	(−1.01)
21	13,114	12,663	1,035.61	248	766	−518
	(6.17)	(1.99)	(4.08)	(1.99)	(6.17)	(−4.18)
22	13,798	12,910	1,068.78	247	661	−414
	(5.22)	(1.95)	(3.20)	(1.95)	(5.22)	(−3.27)
23	14,399	13,159	1,094.23	249	563	−314
	(4.36)	(1.93)	(2.38)	(1.93)	(4.36)	(−2.43)
24	14,934	13,409	1,113.72	250	489	−239
	(3.72)	(1.90)	(1.78)	(1.90)	(3.72)	(−1.82)
25	15,441	13,657	1,130.63	248	455	−207
	(3.39)	(1.85)	(1.52)	(1.85)	(3.39)	(−1.54)
26	15,936	13,904	1,146.14	247	438	−191
	(3.21)	(1.81)	(1.37)	(1.81)	(3.21)	(−1.40)
27	16,452	14,153	1,162.43	249	450	−201
	(3.24)	(1.79)	(1.42)	(1.79)	(3.24)	(−1.45)
28	16,961	14,402	1,177.68	249	437	−188
	(3.09)	(1.76)	(1.31)	(1.76)	(3.09)	(−1.33)
29	17,461	14,650	1,191.87	248	425	−177
	(2.95)	(1.72)	(1.20)	(1.72)	(2.95)	(−1.23)
30	17,953	14,898	1,205.6	248	413	−165
	(2.82)	(1.69)	(1.11)	(1.69)	(2.82)	(−1.13)

Both the capital-labor ratio and decomposition analysis indicate evidence contrary to Fei and Ranis' findings. As shown in Table 1, the capital-labor ratio of Japan had steadily increased over the years under consideration. All of the horizontal effects of the labor absorption equation indicate negative signs, implying that Japan had continuously applied capital-intensive technology throughout her period of development. The findings of this reexamination of Japanese experience are in agreement with those of Shigeru Ishiwata, Ohkawa, Shinora, and Umemura (1966) and T. Watanabe.

IV. Empirical Relevancy of Some of Assumptions and Operations of the Ranis-Fei Model

In addition to these tests, it is of interest to investigate the possibility of verifying the other theoretical assumptions of the Fei-Ranis analysis; such as the relative rigidity of the minimum "institutional" wage in the agricultural sector, and the role of international trade and its implications on the crucial "terms-of-trade effect." Of course, the empirical data that will be examined in this section are the results of Japanese economic performance in the setting as an open economy. Although Fei and Ranis contend that the major conclusions of their model are still intact without the assumption of a closed economy (1964, p. 200), we will see that is not so.

If the theoretical assumptions and operations of the Ranis-Fei model are empirically relevant to historical Japan, we should be able to identify the following trends in its statistical data: 1) a relatively stable trend in real wages of the agricultural and industrial labor at the minimum institutional level in the earlier years before "the commercialization point"; 2) fluctuating terms of trade between the agricultural and industrial sector around parity as a result of the terms-of-trade effects; 3) the two sectoral investments counteracting the terms-of-trade effects; and 4) a trend of meeting "the input-output criteria" in a domestic balance sheet of agricultural commodities without a significant reliance on international trade.

Unfortunately, Japanese historical data do not show these trends conforming to Fei and Ranis' prescriptions. As shown in Table 2, the agricultural and industrial wages did not remain more or less constant at the institutional minimum level throughout the period under consideration, even before the turning point. The agricultural wage level rose as fast as the industrial wage maintaining the Lewisian "cliff" of a wage differential on the order of about 25 percent as shown in column 6 of the table. This trend can be further supported by a general observation of historical Japan by W. W. Lockwood (p. 177): " . . . the food intake of the average Japanese appears to have risen appreciably in forty or fifty years ending in 1930. . . . Rice consumption per capita seems to have risen about 25 percent."

Long-term Japanese price statistics on agricultural and manufacturing products did not show any tendency counteracting the identified over-investiment in the nonagricultural sector. As Table 3 indicates,[5] the

TABLE 2—DAILY MONETARY AND REAL WAGES OF AGRICULTURAL AND MANUFACTURING WORKERS IN JAPAN, 1886–1930
(Five-year Averages)

Year	Agric. Wages (in sen) (1)	Manuf. Wages (in sen) (2)	Consumer Price Index (1934–36 =100) (3)	Agric. Real Wages (in sen) (4)	Manuf. Real Wages (in sen) (5)	(4)/(5) × 100 (6)
1886–90	14.5	20.2	30.78	47.1	65.6	71.8
1891–95	18.0	23.8	31.72	56.7	75.0	75.6
1896–1900	27.0	35.6	43.46	62.1	81.9	75.8
1901–05	32.0	44.2	51.26	62.4	86.2	72.4
1906–10	39.2	56.8	58.54	67.0	97.0	69.1
1911–15	46.2	63.8	62.94	73.4	101.4	72.4
1916–20	87.0	116.0	104.96	82.9	110.5	75.0
1921–25	160.2	201.2	129.28	123.9	155.6	79.0
1926–30	142.8	203.6	117.78	121.2	172.9	70.1

Source: Ohkawa, Shinora, and Umemura, *Prices*, pp. 134, 243, and 245).

---

which of the two quadratic functions for Japan is used in their decomposition analysis (1963, p. 311). In my analysis, the actual incremental ratio of capital overtime, i.e., $\eta_K \equiv (dk/dt)/K$, and that of industrial labor overtime, i.e., $\eta_L \equiv (dL/dt)/L$, are used. Then, the horizontal effect, η_h, is determined as the residual (1963, pp. 285 and 291).

[5] The general direction of Table 3 should not be interpreted as an indication of the movement toward favorable agricultural terms of trade. Note that terms

TABLE 3—TERMS OF TRADE BETWEEN THE AGRICULTURE AND MANUFACTURING SECTORS IN JAPAN, 1886–1930

(Five-year Averages)

Year	Price Index of Agricultural Products (*A*)	Price Index of Manufacturing and Mining Products (*N*)	(*N*)/(*A*)×100
1886–90	24.8	35.2	141.9
1891–95	31.4	38.4	122.3
1896–1900	45.4	54.2	119.4
1901–05	51.4	62.9	122.4
1906–10	58.4	75.4	129.1
1911–15	67.2	78.0	116.1
1916–20	121.3	153.1	126.2
1921–25	141.4	150.4	106.4
1926–30	111.3	118.3	106.3

Source: Ohkawa, Shinora, and Umemura, *Prices*, pp. 165 and 192–93; (1934–36=100).

TABLE 4—DEMAND AND SUPPLY OF AGRICULTURAL PRODUCTS IN JAPAN, 1878–1917

(Five-year Averages)

Year	Domestic Supply	Domestic Demand	Supply—Demand
1878–82	431	420	11
1883–87	333	326	7
1888–92	460	463	−3
1893–97	632	663	−31
1898–1902	959	1,058	−99
1903–07	1,230	1,425	−195
1908–12	1,501	1,754	−253
1913–17	1,787	2,186	−399

Source: Ohkawa and Rosovsky, p. 47; *Unit*: Million Yen.

agricultural sector had continuously suffered unfavorable terms of trade as compared to the industrial sector in spite of the chronic shortages of agricultural products that began as early as 1880's (see Table 4). Even the relative changes in the level of agricultural prices are not consistent with the Fei-Ranis thesis.[6]

Furthermore, it is clearly apparent that the terms-of-trade effects did not have any significant role in resource allocation and in restoring the balanced-growth path in view of historical Japanese sectoral capital formation. The coexistence of chronic shortages of agricultural products, consistently unfavorable agricultural terms of trade and continuous heavy investments in the industrial sector empirically nullifies and is paradoxical to Fei and Ranis' terms-of-trade effects. This apparent paradox may be explained partly by a remarkable increase in the agricultural productivity in the early stage of Japanese development as documented by Ohkawa and Rosovsky and more significantly by the role of international trade, i.e., importation of agricultural products from Korea and Taiwan, as conjectured by Oshima, and as empirically seen in Table 4. In this context, it is doubtful whether or not the conclusions from the Ranis-Fei model would be intact even when the foreign sector is a "facilitator" of growth.

V. Conclusions

Based on the preceeding examination of empirical evidence of historical Japan, the conclusions must be that the data do not support the Ranis-Fei model. In contrast to the Ranis-Fei model, historical Japanese development is clearly characterized by a persistent capital deepening trend, by consistent use of capital-intensive technology, and the unbalanced-growth path in their sense. There was no evidence to identify the alleged rigidity of the institutional minimum wages nor the terms-of-trade effects interacting upon the two sectors and to restore the balanced-growth path. The evidence further suggests the weakness of the concept of disguised unemployment as noted by Oshima. On the other hand, foreign trade did have a vital role in supplementing the deficiency in food supply and promoting Japanese development.

The Ranis-Fei model may be theoretically consistent, but it is not empirically relevant. Thus, Fei and Ranis might have succeeded in emphatically advancing "the point of view of the logical necessity of capital shallowing and its desirability" (1964b, p. 1063),

---

of trade of the two-sector commodities are expressed as relatives of price indices. Therefore, the relative of indices should converge to the parity as it approaches the base year, 1934–36.

[6] This point was indicated to the writer by the unknown referee of this comment.

given the assumptions. However, "its empirical validity in the case of historical Japan" (1964b, p. 1063) has no ground. Consequently, serious doubt must be raised as to the validity of their investment criterion, balanced-growth path, and the compatibility of the maximum-employment and the maximum-output, at least for empirical reasons.

It is also evident that to brand the case of Indian experience as a contrast to a "successful" case is a hasty judgement, not only for the reason that the period of Indian experience considered is too short,[7] but for the very reasons that historical Japan was successful.

Problems of economic development are indeed complex and diversified. Therefore, in any effort to prescribe a historical uniformity as a model, such complexity and diversity should receive due considerations. And in this "quantitative age," abstraction of facts from figures and empirical support of a model should be done with extreme care avoiding possible "sins of omission."

REFERENCES

R. S. Eckaus, Book review of *Development of the Labor Surplus Economy—Theory and Policy*, by Fei and Ranis; *Amer. Econ. Rev.*, Mar. 1965, *55*, 201–04.

J. C. H. Fei and G. Ranis, "Capital Accumulation and Economic Development," *Amer. Econ. Rev.*, June 1963, *53*, 283–313.

——— and ——— (1964a) *Development of the Labor Surplus Economy—Theory and Policy*, Homewood 1964.

——— and ——— (1964b) "Capital-Labor Ratios in Theory and in History: Reply," *Amer. Econ. Rev.*, Dec. 1964, *54*, 1063–69.

U. Datta and V. Prakash, "An Estimate of Reproducible Tangible Wealth in India," *Papers on National Income and Allied Topics*, Vol. 1, New York 1960.

S. Ishiwata, "Capital Stock in Japan, 1878–1940," *J. Soc. Sci.* (Japanese), 1965, *6*, 243–68.

M. Ito, "Estimates of Capital Coefficient and Rate of Savings by National Wealth Statistics," in K. Ohkawa et al., *The Growth Rate of the Japanese Economy since 1878*, Tokyo 1957.

W. A. Lewis, "Economic Development with Unlimited Supplies of Labour," *Manchester Sch. Econ. Soc. Stud.*, May 1954, *22*, 139–91.

W. W. Lockwood, "Scale of Economic Growth in Japan, 1868–1938," in S. S. Kuznets, W. E. Moore, and J. J. Spengler, eds., *Economic Growth: Brazil, India, Japan*, Durham 1955.

K. Ohkawa et al., *The Gowth Rate of the Japanese Economy Since 1878*, Tokyo 1957.

K. Ohkawa and H. Rosovsky, "The Role of Ariculture in Modern Japanese Economic Development," *Econ. Develop. Cult. Change*, Oct. 1960, no. 1, pt. 2, 43–67.

K. Ohkawa, M. Shinohara, and M. Umemura, *Estimates of Long-Term Economic Statistics of Japan Since 1868*, Vol. 3, *Capital Stock*, Tokyo 1966 (Japanese, summary in English).

———, ———, and ———, *Estimates of Long-Term Economic Statistics of Japan Since 1868*, Vol. 8, *Prices*, Tokyo 1967 (Japanese, summary in English).

H. T. Oshima, "The Ranis-Fei Model of Economic Development: Comment," *Amer. Econ. Rev.*, June 1963, *53*, 448–52.

G. Ranis, "Factor Proportions in Japanese Development," *Amer. Econ. Rev.*, Sept. 1957, *47*, 594–607.

——— and **J. C. H. Fei**, "A Theory of Economic Development," *Amer. Econ. Rev.*, Sept. 1961, *51*, 533–58.

——— and ———, "The Ranis-Fei Model of Economic Development: Reply," *Amer. Econ. Rev.*, June 1963, *53*, 452–54.

E. P. Reubens, "Capital-Labor Ratios in Theory and in History: Comment," *Amer. Econ. Rev.*, Dec. 1964, *54*, 1052–62.

H. Rosovsky, *Capital Formation in Japan, 1868–1940*, New York 1961.

W. W. Rostow, *The Stages of Economic Growth: A Non-Communist Manifesto*, Cambridge 1960

T. Watanabe, "Economic Aspects of Dualism in the Industrial Development of Japan," *Econ. Develop. Cult. Change*, Apr. 1965, *13*, 293–312

---

[7] See the book review by R. S. Eckaus, p. 203.

On the Empirical Relevancy of the Ranis-Fei Model of Economic Development: Reply

By JOHN C. H. FEI AND GUSTAV RANIS*

The attempt to construct an analytical model aimed at explaining some crucial facets of economic history which can be statistically verified is clearly a worthwhile experiment in growth theoretic methodology. Our own, admittedly crude, attempt to accomplish this for the case of Japan (1880–1930) with the help of imperfect data has been scrutinized by a number of critics (see E. P. Reubens, H. Oshima, D. Jorgenson, and Hakchung J. Choo).[1] We are encouraged by the present increased interest in using the historical laboratory to obtain a better understanding of development and welcome the opportunity to further clarify the issues raised.

The type of model under discussion is primarily a growth model which attempts to depict, in broad outline, the *modus operandi* of a particular type of economy over a particular stretch of its history, i.e., in the specific case before us, the development of a labor surplus economy, Japan, between 1880 and 1930. As is necessary in constructing any model, we chose to make certain sharp simplifying assumptions, which, in turn, led to a number of sharp deductions. One of our main assumptions for example, was the near constancy of the institutional real wage; from this we were able to deduce capital shallowing tendencies in the nonagricultural sector.

What Choo emphasizes is that he does not find constancy in the real wage, nor "pronounced" capital shallowing in the nonagricultural sector for the period under discussion. Admittedly his 1968 data sources were not available to us in 1964; we could, moreover, quibble about the fact that more recent revised data of our own do show capital shallowing up to 1887. But we do not wish to debate in terms of who has the better data; for the sake of argument, therefore, we accept all of Choo's data in this reply. The question we wish to address, instead, is whether or not his evidence really contradicts, as he claims, the basic theme we developed in interpreting Japanese growth. We believe it does not.

The main theme we attempted to portray is that a labor surplus economy can learn only gradually to make full use of its potentially most valuable resource, i.e., its surplus labor, in the course of development. The successful exhaustion of that surplus over a forty-year period (1885–1925), in fact, constitutes the core of our interpretation of Japanese economic development. In any rational society we would expect such a process to be accompanied by (i) slow increases, if any, in the real wage, as long as surplus labor overhangs the market,[2] giving way to more rapid increases once that surplus is exhausted; (ii) little capital deepening (or even, perhaps, capital shallowing) in the first phase, giving way later to rapid capital deepening, again reflecting the advent of labor scarcity as "success" is achieved. In historical perspective, in other words, our central thesis can be supported or rejected in a really meaningful sense only in terms of the existence or non-existence of two such distinct historical phases of growth.

Let us try to understand these two phases a little more fully. The second condition (ii) above can be deduced from the first (i) by reference to assumptions of rational entrepreneurial behavior. Even in the absence of technological change, i.e., with fixed pro-

---

* Professors of economics, Yale University.

[1] See also our replies to earlier comments (1964, pp. 1063–69; 1963, pp. 452–54; 1966).

[2] The fact that the wage increases at all at this time—a possibility mentioned but not emphasized in our previous work—is due to the fact that, as average labor productivity in agriculture grows, it is natural to expect the institutional wage to experience some secular upward drift, and, similarly, with respect to the industrial wage at a margin above the agricultural wage.

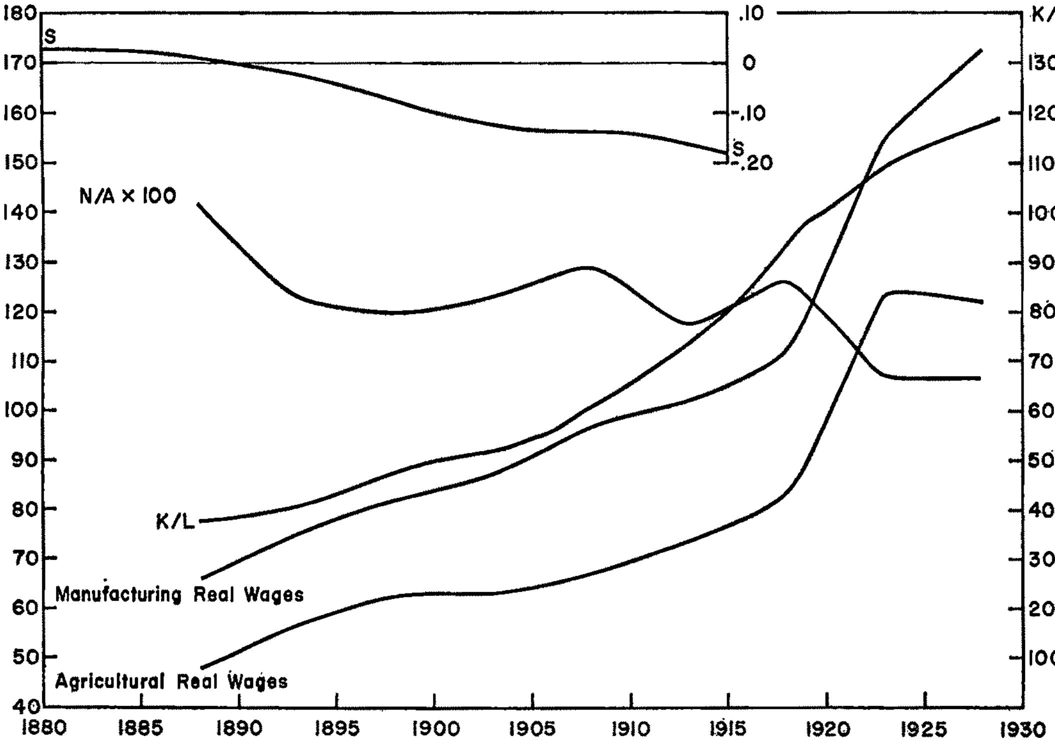


FIGURE 1

duction functions, (ii) is already a direct consequence of (i) as a result of profit maximization. Superimposed on this simple causal relationship is innovational behavior which causes a tendency toward capital shallowing (or relatively less capital deepening) when the increase in the real wage is slow (or the wage is nearly constant), and a tendency toward capital deepening when wage increases are fast, at a later stage. Thus, in the first phase, if wages are, in fact, rising somewhat, capital shallowing may not be observable—Choo's main contention—unless the labor-using bias of innovations is strong enough to compensate for the moderate capital deepening effect due to the moderate wage increase.[3]

Using Choo's data, we can plot the time series for K/L from his Table 1 in our Figure 1. Choo stresses only the continuously increasing trend in this ratio, i.e., sustained capital deepening, and omits all consideration of possible differences in different phases of growth. Inspection of the figure, however, indicates the existence of a significant turning point around 1906,[4] marking off two phases with substantially different average annual rates of capital deepening ($\eta_{K/L}$):

1880–1906	1906–1930
2.1%	3.3%

The significantly higher value in the second phase (i.e., higher by 50 percent) appears to support our central thesis. To probe the matter further, let us compute the values of $\eta_{K/L}$ for the following shorter sub-phases:

1892–1900	1900–1906	1906–1917	1917–1930
2.8%	1.7%	4.3%	2.4%

The point to be stressed here is the significantly lower value (1.7 percent) during the second sub-phase. This seems to suggest that, before the turning point in 1906, Japanese entrepreneurs were getting better and better, through a learning-by-doing process, at innovating in a relatively labor-using direction. Moreover, after 1906 we note a marked upward revision, indicating the switch of innovations in a labor saving direction after the turning point.[5] All this is in perfect accord with the central theme of our analysis of Japanese growth. It is, in fact, historical evidence of this type that led to our counsel to the contemporary developing society, i.e., India, with respect to the desirability, if not the necessity, of undergoing a phase of relative "capital shallowing" in order to make fuller use of her surplus labor. Surely the fact, cited by Choo, that India in 1949 had a less favorable resource endowment than Japan in 1888, argues for even greater efforts to utilize her relatively abundant labor supply rather than engaging in a search for even greater capital intensity.[6]

---

[3] More technically put, even if innovations are "not very labor-saving," some net capital deepening may result from the wage increase.

[4] This is admittedly somewhat earlier than the turning point arrived at with the data available in 1964, but coincides with the earlier findings of one of the authors (Ranis, 1957) cited by Choo.

Time-series on agricultural and industrial real wages, taken from Choo's Table 2, are also shown in our Figure 1. They yield the following annual average rates of growth of wages in the two sectors:

	1880–1908	1908–1928
agricultural wage η_{w_a}	1.7%	2.7%
industrial wage η_{w_i}	2.0%	2.9%
difference	.3%	.2%

Thus, whether we are considering the agricultural or the industrial real wage,[7] this permits us to conclude that the rate of increase is significantly larger after than before the same turning point—a finding which clearly supports our basic hypothesis. To probe this matter a little further, let us calculate the rates of growth of real wages in both sectors by shorter sub-phases (rows 1 and 2):

	1880–1898	1898–1918	1918–1928
(1) agricultural wage (w_a)	2.8%	1.4%	3.7%
(2) industrial wage (w_i)	2.2%	1.4%	4.4%
(3) terms of trade	−1.6%	0%	−1.6%

This permits us to note that a relatively moderate increase in both wages during the twenty-year period (1898–1918) gives way to relatively sharp increases after World War I, with the exhaustion of the labor surplus. We can thus clearly see a differentiable pattern before and after the turning point concerning the behavior of real wages and the nature of technological change outside of agriculture.

Choo also claims that our balanced-growth thesis, an essential part of the explanation of successful development in a dualistic economy, is defective. The modernization of agriculture at just the "right" speed is certainly a complicated phenomenon. Any adequate analysis of it must take into consideration at least the following three facets: incentives to private cultivators, emphasized formally in our work through the terms of trade argument; government policies, dealt with informally; and the opportunity of foreign trade, emphasized by Choo. All these factors must interact in such a way that, as surplus labor is drawn out of the agricultural sector, increases in agricultural productivity, plus imports, are just enough to provide food and raw materials for the newly reallocated workers. A slackening of agricultural productivity increase will cause a food shortage and an increase in the industrial real wage through a worsening of the industrial sector's terms of trade; and vica versa for the case of rapidly expanding agricultural productivity (see Fei and Ranis (1964, ch. 5)).

While the story of modernization of agriculture is indeed complicated, the fact that Japan is a relatively land scarce country permits us to make two observations on a priori grounds. First, the success story of Japanese agriculture must at least have been encouraged by a long-run favorable trend in the terms of trade to permit proper entrepreneurial and technological responses on the intensive margin. Second, if such a country wants to grow quickly, it may, in spite of successful agricultural modernization, have to resort increasingly to food imports—especially as the initial waves of innovation through diffusion of better practice are exhausted (see Y. Hayami and S. Yamada).

---

[5] We think we can explain the decline of the rate of capital deepening after a war period (World War I), but do not consider this factor relevant to the present controversy.

[6] Choo's relationship between late-comer advantages and high capital intensity escapes us altogether.

[7] Notice that the wage multiple w_i/w_a maintains the same annual rate of increase (between .2% and .3%) for the entire period—as can also be seen from Figure 1. This lends further support to the surplus labor theory which has the industrial wage "tied" to the agricultural wage.

To pursue this further, let us plot Choo's terms-of-trade data (his Table 3) once again in our Figure 1, i.e., the N/A curve. The indicated behavior of the terms of trade can now be divided into three phases: (i) an early trend (1888–1898) of improvement in favor of agriculture; (ii) approximately twenty years (1898–1918) of minor fluctuations around a nearly constant trend; (iii) a resumption of the movement favorable to agriculture after World War I (1918–1928). This can also be demonstrated by computing the average annual rates of growth of N/A, presented in row three (3) of the last table (above).

Two points should be noted. First, there is, indeed, evidence of a long-run movement of the terms of trade in favor of the agricultural sector—contrary, incidentally, to the conclusion Choo draws from his own data. Second, the first and last sub-phases, in which this tendency is most marked, are also precisely those in which the industrial real wage, as we saw earlier, grows relatively faster—confirming our basic hypothesis.

The increasing reliance over time on imported agricultural goods to meet domestic food demand is indicated in Choo's Table 4 and depicted in our Figure 1 by the SS curve—representing the degree of self-sufficiency, i.e., the percentage of total food demand produced domestically. Notice that a food export surplus gives way to an import surplus around 1890, permitting Japan to forestall any further deterioration of her industrial sector's terms of trade and consequently greatly lessening the upward pressure on industrial wages during the next twenty years. This increasing reliance on imports to make up for the slackening performance of a domestic agriculture no longer favored by improving terms of trade contributed to the postwar crisis symbolized by the Rice Riot of the early 1920's. Japan responded by assigning increasingly important food producing roles to her colonies, mainly Taiwan and Korea (see Yujiro Hayami and Vernon W. Ruttan).

At the same time, the terms of trade were permitted to move once again in favor of agriculture, providing a stimulus to agricultural investment and technological change. But the observed sharp increase in the industrial real wage after World War I was much more due to the exhaustion of the reservoir of surplus labor in agriculture, combined with the introduction of industrial minimum wage legislation attendant on Japan's membership in the ILO.

In summary, the purpose of our framework of analysis was to explain how a dualistic labor surplus economy can convert its excess manpower into a resource for development via the balanced expansion of its agricultural and nonagricultural sectors. No model of economic history is self-contained—or can be expected to fully replicate an always too complicated reality, leaving much still to be "explained" by exogenous factors. Nevertheless, a useful growth model can and should help us gain a coherent interpretation of the main thrust of historical experience in bare outline form. We certainly do not claim more. To evaluate such a model, the empirical evidence should be read not in terms of a narrow interpretation based on the model's simplifying assumptions but on whether or not the essence of the two stages thesis—and the different rules of behavior obtaining within each—is corroborated.

REFERENCES

J. C. H. Fei and G. Ranis, "Agrarianism, Dualism and Economic Development," in I. Adelman and E. Thorbecke, eds., *The Theory and Design of Economic Development,* Baltimore 1966.

——— and ———, "Capital-Labor Ratios in Theory and in History: Reply," *Amer. Econ. Rev.,* Dec. 1964, *54,* 1063–69.

——— and ———, *Development of the Labor*

Surplus Economy—Theory and Policy, Homewood 1964.

——— and ———, "The Ranis-Fei Model of Economic Development: Reply," *Amer. Econ. Rev.*, June 1963, *53*, 452–54.

Y. Hayami and V. W. Ruttan, "Korean Rice, Taiwan Rice, and Japanese Agricultural Stagnation: An Economic Consequence of Colonialism," *Quart. J. Econ.*, Nov. 1970, *84*, 562–89.

Y. Hayami and S. Yamada, "Technological Progress in Agriculture," in L. Klein and K. Ohkawa, eds., *Economic Growth: The Japanese Experience Since The Meiji Era,* Homewood 1968.

D. Jorgenson, "Testing Alternative Theories of the Development of a Dual Economy," in I. Adelman and E. Thorbecke, eds., *The Theory and Design of Economic Development,* Baltimore 1966.

H. Oshima, "The Ranis-Fei Model of Economic Development: Comment," *Amer. Econ. Rev.*, June 1963, *53*, 448–52.

G. Ranis, "Factor Proportions in Japanese Development," *Amer. Econ. Rev.* Sept. 1957, *47*, 594–607.

E. P. Reubens, "Capital-Labor Ratios in Theory and in History: Comment," *Amer. Econ. Rev.*, Dec. 1964, *54*, 1052–62.

Incentive and Disincentive Experimentation for Income Maintenance Policy Purposes: Note

By EDGAR K. BROWNING*

In a recent article in this *Review*, Guy H. Orcutt and Alice G. Orcutt discuss the use of experimental techniques in predicting the quantitative outcomes of income maintenance policies. The experimental procedure recommended is to take a random sample from the entire group which would be affected by the policy under consideration. The sample is to be subdivided into two sub-samples, the first to receive negative income tax transfers and the second to act as a control group and receive no such transfers. Orcutt and Orcutt assert that the results of such a study provide the basis for an unbiased estimate of the magnitude of the disincentive effects of the policy under consideration:

> From the first sample we could obtain an unbiased estimate of what the earned income of the entire group would have been if it had been exposed to a negative income tax. From the second sample we could get an unbiased estimate of what the earned income of the same group for the same time period would have been if the entire group had not been exposed to the negative income tax. [p. 758]

The purpose of this note is to show that this assertion is incorrect—that the results of such an experiment cannot be generalized to provide reliable information concerning the disincentive effects of an income maintenance policy. The reason is that the aggregate market effects of such a policy will not occur in the sample study, and cannot be inferred from the experimental results. To illustrate the argument, I will assume that a specific negative income tax (*NIT*) plan is the policy under consideration. In addition, it will be assumed that there are no other welfare programs in existence, so that additional problems relating to the substitution of a *NIT* for other programs will not be encountered.

In Section I, I analyze briefly the effects of the *NIT* policy implemented on a national basis in a competitive economy where individuals are free to vary their work effort at the market wage rate. These effects are contrasted with the predictions of an experimental study of the type outlined above. In Section II, I allow for institutional rigidities in labor markets, and show how these will affect the outcomes of a national *NIT* policy and of an experimental study.

I. Aggregate Market Effects

Let *ON* in Figure 1 be the budget constraint relating income to hours worked for a household. The *NIT* shifts the constraint to *MBN* where *OM* is the income guarantee and *B* is at the breakeven level of income. The initial effect of this is that the household moves from *A* to *C*, reducing work effort from H_1 to H_3. Both the income effect

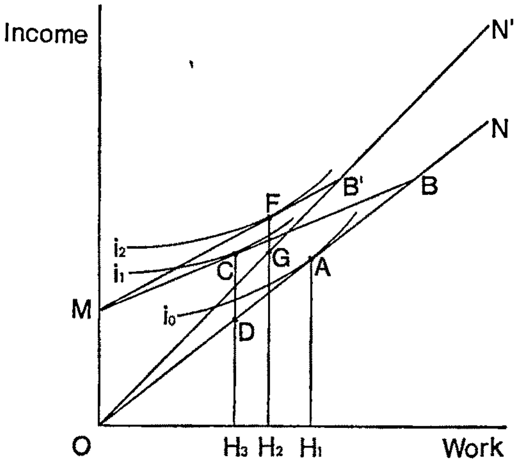


FIGURE 1

---

* Acting assistant professor of economics at the University of Virginia. I am grateful to Roland N. McKean, Roger P. Sherman, and Ivan C. Johnson for comments and suggestions on an earlier draft.

(following a higher income at each level of work effort over the relevant range and assuming leisure is not inferior) and the substitution effect (resulting from a lower net wage rate) encourage less work. All households with an income less than the break-even level will have a similar incentive to reduce their work effort.[1]

This initial reduction in work effort on the part of the transfer recipients will show itself in labor markets as a reduction in the supply of labor. In Figure 2, the *DD* and *SS* curves represent the aggregate demand and supply curves for the type of labor which will receive *NIT* transfers. Prior to implementation of the *NIT*, W_1 is the market wage rate (which can best be thought of as a proxy for the level of wage rates generally), L_1 is the quantity of labor supplied and demanded, and OW_1KL_1 is the earned income of these workers. The *NIT* produces a shift in the supply curve to $S'S'$. The market wage rises to W_2, the equilibrium quantity of labor supplied falls to L_2, and earned income becomes OW_2JL_2. Earned income may be greater or less than formerly, depending on whether the elasticity of demand is smaller or larger than unity.

The impact of this change in the market wage rate on the individual household's budget constraint is illustrated in Figure 1. The straight line ON' becomes the new constraint relating earned income and hours of work, and $MB'N'$ is the new budget constraint including the *NIT* transfer. (Note that B' occurs at the same level of earned income as B: the break-even level, which is defined as the level of earned income at which no transfer is received, has not been changed.) Suppose that the household moves from C to F, increasing work effort from H_3 to H_2 and earned income from DH_3 to GH_2.[2] Work effort will, of course, be reduced below H_3 in response to the higher market wage rate if the supply curve is backward bending. Figure 1 is drawn with an increase to H_2 to conform to the upward sloping supply curve in Figure 2.

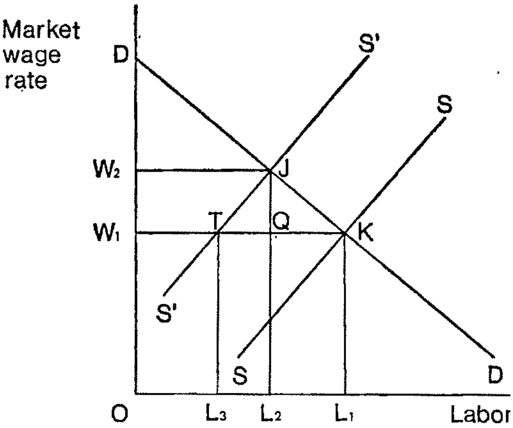


FIGURE 2

This, in perhaps overly concise form, is the outcome that would be predicted from standard economic theory. Now let us consider the predictions which an experimental study of this *NIT* plan would generate. From among those initially supplying the L_1 units of labor, a random sample is drawn and subjected to the payment of *NIT* transfers. Suppose that our representative household in Figure 1 is selected for this treatment. Then its budget constraing shifts to *MBN*, and it moves to point *C*.

And this is all that happens. For the sample will always be a negligible part of the entire market, so that there will be no noticeable pressure on the market wage rate to change.

Given that the sample is random, and that there are no "Hawthorne"-type problems,[3] the experiment would predict a reduction in the quantity of labor supplied from L_1 to L_3 in Figure 2. This corresponds to the reduction from H_1 to H_3 in Figure 1. What would be estimated, in fact, is the amount by which the quantity of labor supplied would fall at the existing market wage rate, i.e., the horizontal shift in the supply curve. It would also be predicted that earned income would fall to OW_1TL_3.

---

[1] It is often the case that households with incomes above the break-even level will have an incentive to reduce work effort also, but this point will be ignored.

[2] The shift in the supply curve can be illustrated in Figure 1 by drawing in the two price consumption curves. The price consumption curve when the *NIT* is in effect lies to the left (at lower work effort) of the original curve at each market wage rate, at least until the household earns an income above the break-even level.

[3] For an explanation and discussion of the "Hawthorne Effect," see Orcutt and Orcutt, pp. 759–61.

Now we can easily contrast the experimental predictions with the actual outcome. Referring to Figure 2 again, the prediction regarding work effort is a fall to L_3. In actuality, work effort may be greater or less than L_3, following enactment of a national *NIT* program. It will be greater if the supply curve is upward sloping through T (as it is drawn) and it will be less if the supply curve is backward bending.[4] The prediction regarding earned income is a fall to OW_1TL_3. In fact, earned income may not fall from its initial level at all: if the demand curve is inelastic, earned income will rise following enactment of the *NIT* on a national basis, as noted earlier. In any event, earned income will always be greater than OW_1TL_3,[5] since the change in the market wage rate produces a movement up the new supply curve from point T, and earned income always increases with a move up a supply curve, whatever its slope.

In addition, the experiment would provide a prediction of the fall in national product of TKL_1L_3, the estimated reduction in labor supply weighted by the initial market wage rate. The actual reduction in Figure 2 is JKL_1L_2: whether this is greater or less than the predicted reduction depends on (in general) the elasticities of both the demand and the supply curves.

A final point might be noted. A *NIT* will generally change the rate of return for other factors of production (which include more highly paid labor which does not receive *NIT* transfer payments). If we assume that the other factors are in completely inelastic supply,[6] then the income of these factors will fall from DW_1K to DW_2J, or by W_1W_2JK.[7,8] This is simply because the marginal productivity of these factors falls when the quantity of the factor they are combined with in production (labor receiving *NIT* transfers) is reduced. An experiment will provide no way to estimate this effect on the rate of return for other factors of production.

Thus, in this world where individuals are free to vary work effort, an experimental study will not provide unbiased estimates of the effects of a *NIT* policy. The reason is simply that the experiment will, at best, just estimate the horizontal shift in the supply curve, but will give no indication of the consequences of this shift—which is what we would really like to know.

It would obviously be helpful to know how much the supply curve will shift. It might be argued that if the shift is relatively small, then the aggregate market repercussions discussed here will be of little importance. In the next section, however, I will argue that it is doubtful that an experiment will provide even this piece of information.

II. Group Versus Individual Adjustment

In the last section it was assumed that individual households could vary freely the number of hours worked. However, in reality, individual households cannot, in general, work any number of hours they please. There are well-known institutional rigidities in labor markets which preclude this possibility. These rigidities usually take the form of all-or-nothing offers from employers. For example, an employer will offer a wage rate contingent on the employee working a forty-hour week. The employee must perform forty hours of work per week, or else he will not for long hold the job.

Taking account of such institutional rigidities does not alter the effects that would be expected from a national *NIT* program. Whereas no single individual may be able to influence (say) the length of the work week, the employers have incentive to tailor their offers to the desires of the entire group of workers. Since the *NIT* will result in a general desire for shorter hours of work, the institutional rigidities will break down as employers compete for workers in markets where the aggregate supply conditions have changed. For this reason, institutional rigidities such as a fixed work week must be considered as variables in the analysis of a general change in the desires of workers.

---

[4] If the experiment is testing a number of *NIT* plans with different marginal tax rates, it may give some information about the elasticity of the supply curve. However, this possibility is weakened by the consideration raised in Section II.

[5] If the experimental results are used to estimate the total government outlay on the *NIT*, then an underestimate of earned income means that the outlay will be overestimated since the transfers vary inversely with earned income under a *NIT* arrangement.

[6] This has actually been implicit all along. If the quantity of other factors changes, then the demand curve in Figure 2 will shift.

[7] This, of course, does not show the additional fall in income resulting from the higher taxes necessary to finance the program.

[8] This does not mean that the rate of return for each and every factor will fall. Factors which are very good substitutes for the type of labor receiving *NIT* transfers may rise in price.

These institutional rigidities, however, will not be affected by an experimental study. The households in the sample may be unable to adjust their work effort to a lower level when the simulated *NIT* encourages them to do so because the restrictions on the nature of job offers are beyond the control of any single household or small group of households. Only when a truly general desire for shorter hours arises will employers have sufficient incentive to change their offers. An experimental study will produce no such general desire, but only a desire on the part of the households in the sample.

If this is correct, then an experiment of the type discussed will not provide an accurate prediction of the shift in the supply curve. What such an experiment will provide is information about the responses of a small number of households to the *NIT*, given that the institutional elements in the labor market are beyond their control. These institutional elements are *data* for an experimental study, but they are *variables* for a national policy. Thus there is no reason to expect an equivalence in the responses of households in these two situations.

The assertion that households will respond to an experiment in the same way as they would to a national policy (which seems to be implicit in the article by Orcutt and Orcutt) reflects a misunderstanding of the nature of the choices open to households. The income-leisure apparatus traditionally employed in analyzing work effort responses is only useful in examining general changes in preferred levels of work effort, such as would result from a *NIT* policy of wide coverage. It is not very useful in examinin how a single household (or a small numb of households) will respond to a policy whe it is the only one subjected to this polic The opportunities open to a household a simply not the same in these two situations In the former case, it is legitimate to procee with the analysis *as if* each household coul freely choose its amount of work effort, f employers are given incentive to adju their offers to the preferences of workers. I the latter case, this is not a legitimate ste because the members of a sample are r stricted in their choices by the unchanged d sires of all the other workers in the marke

In reality, there is, of course, some scop for individual variation in work effort. Th wider the range of choices open to *indiv duals*, the more likely it is that an experimen would provide a clue to the actual shift in th supply curve which a policy of nation coverage would produce. I have no idea ho much individual variation would be possib for the participants in an experimental stud The only point I wish to make is that th variation possible when a national policy implemented is surely greater than tha possible for a participant in such a stud This means that an experiment cannot ge erate an unbiased estimate of the shift i supply which would be produced by a na tional policy.

REFERENCES

M. Friedman, *Price Theory: A Provision Text*, Chicago 1962.

G. H. Orcutt and A. G. Orcutt, "Incenti and Disincentive Experimentation for I come Maintenance Policy Purposes," *Ame Econ. Rev.*, Sept. 1968, *58*, 754–72.

---

[9] The important distinction between the situati from the viewpoint of the individual and from the vie point of the group is concisely pointed out by Milt Friedman: "The important point is that the individu is like the perfect competitor: to each individual sep rately, the number of hours of work per week may fixed, yet the level at which it is fixed is the result of t choices of the individuals as a group" (p. 205).

Pecuniary Externalities: A Game Theoretic Analysis

By MARTIN SHUBIK*

When one firm or an industry expands, its actions may influence the cost of factors to another firm. The change in the cost of factors represents an externality to the firm which finds its input prices changed. The firm is not the master of its own destiny. An action by an outside agent has influenced its economic well being.

I. Pecuniary Externalities and Markets

Why should a pecuniary externality, i.e., an externality that is manifested in a change in prices, be different from a physical externality such as the production of smoke or other pollution? Modern economic analysis provides the answer that the change in prices is merely a reflection of the maximization process in a competitive market embodied in a general equilibrium system. The apparent externality shows up as a feedback from one market to another in a comparative statics explanation of adjustment to change.

The paradox of the pecuniary externality is caused by the nature of the partial equilibrium analysis of the firm. A real or a pecuniary externality appear to be the same when viewed from the viewpoint of the individual firm. When embedded in a general equilibrium context the differences emerge.

In our attempt to set up the appropriate economic model we find that the paradox of the pecuniary externality is closely related to the implicit assumption of the existence of markets rather than the assumption of the emergence of markets as part of the economic process.

The paradox does not even appear if, instead of viewing the behavior of the individual firm or customer as a mere mechanistic adjustment to a price system, it is viewed as the acts of a player in an economic game.

Instead of assuming the existence of a price system, we deduce its existence as the limit of the core of an economy with many players (see Shubik). Although the concept of the *core* has gained wide recognition in the past few years, it may still be unfamiliar to some readers. For this reason a definition is supplied in the footnotes.[1]

F. Y. Edgeworth first suggested approaching the formation of the price system via the contracting and recontracting of groups in a market. He indicated the limits of a bargain between traders by means of a contract curve and indicated that as the number of traders is increased the contract curve shrinks and in the limit (with many traders) the price system emerges. In my 1959 article, I noted that the Edgeworth contract curve was the same as the game theoretic concept of the core of a game. I treated a special case; Herbert Scarf treated the general case and proved the convergence of the core to the competitive equilibrium price in a general context. Gerard Debreu worked with Scarf in shortening and strengthening the proof.

The reason for the paradox of the pecuniary externality and why it does not appear when we approach the price system via the core can be explained by means of a simple

---

* Professor, economics of organization, Yale University, on leave to the RAND Corporation. Research undertaken by the Cowles Commission for Research in Economics under Contract Nonr-3055(01) with the Office of Naval Research.

[1] The core is composed of that set of imputations of wealth on the Pareto optimal surface which cannot be blocked by any coalition. An imputation can be blocked if there exists a coalition that could obtain more for all of its members without cooperating with the others. The following simple game illustrates the core. Suppose players 1, 2, 3 can together obtain \$30. The Pareto optimal surface consists of a set of imputations of the form (a_1, a_2, a_3) where $\sum_{i=1}^{3} a_i = 30$. Suppose that any individual alone can obtain nothing and that any pair can obtain \$15. The core consists of the imputations satisfying the conditions $a_1 \geq 0$, $a_2 \geq 0$, $a_3 \geq 0$, $a_1 + a_2 \geq 15$, $a_1 + a_3 \geq 15$, $a_2 + a_3 \geq 15$, $a_1 + a_2 + a_3 \geq 30$, and $\sum_{i=1}^{3} a_i = 30$. It is easy to observe, for example, that the imputations (8, 9, 13) or (10, 10, 10) are in the core.

Not all games have cores. In the example above, suppose that any pair could obtain \$25 then three of the inequalities would become $a_1 + a_2 \geq 25$, $a_1 + a_3 \geq 25$, $a_2 + a_3 \geq 25$ and no set of (a_1, a_2, a_3) adding to 30 can exist.

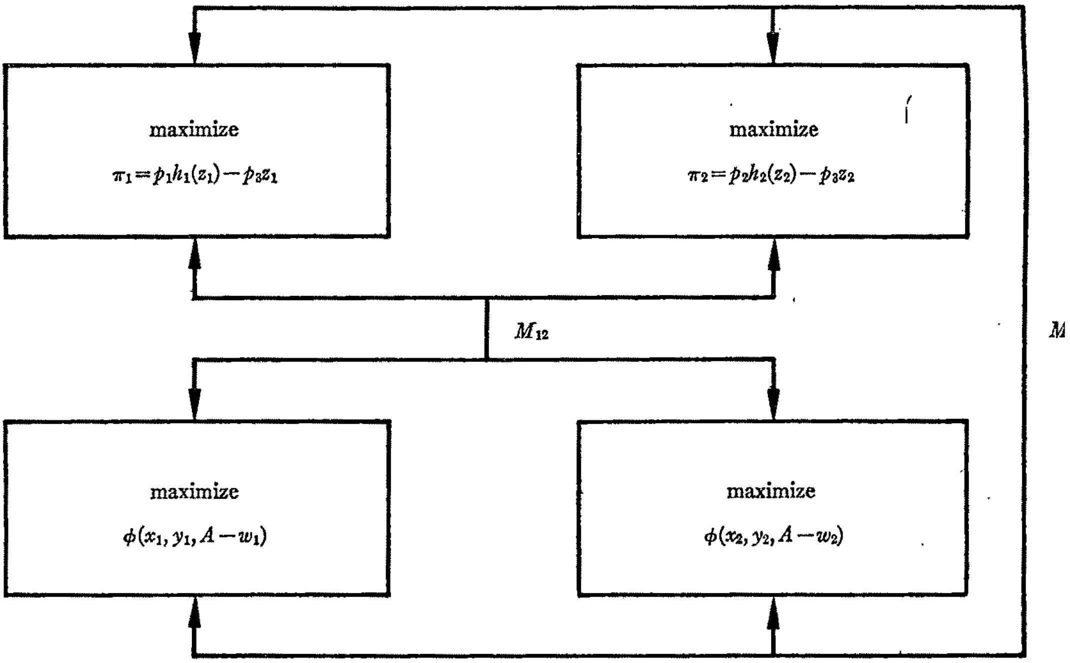


FIGURE 1

example of a closed economy with two industries, customers, two final products, and labor.

Figure 1 shows the general equilibrium market model and Figure 2 shows a closed economy modeled as a game. They are modeled for two firms and two customers. For simplicity, we assume that raw materials are free inputs, hence the profits of firms 1 and 2 can be expressed as:

$$\Pi_1 = p_1h_1(z_1) - p_3z_1$$

and

$$\Pi_2 = p_2h_2(z_2) - p_3z_2$$

where $x=h_1(z_1)$ and $y=h_2(z_2)$ are the production functions for the two consume goods. The distributions of the consume goods to the customers are $x_1+x_2=x$ an $y_1+y_2=y$, where the subscripts to x and

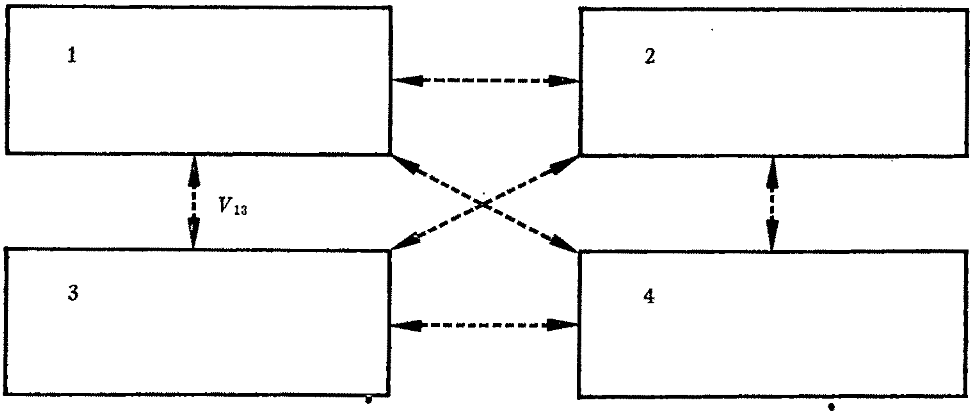


FIGURE 2

denote customers. The amount of labor bought equals the amount sold, i.e., $z_1+z_2=w_1+w_2$. The prices of the first and second consumer goods are p_1 and p_2; the price of labor is p_3. For simplicity we assume that the utility functions of the consumers are the same, ϕ, and their initial endowments are also the same $(0, 0, A)$.

The lines and arrows marked by M_{12} show the presence of a market for the consumer goods (or two interlinked markets). The market structure aggregates and interlinks firms and customers so that any individual trades with an impersonal market rather than indulges in personal face-to-face bargaining and trading. Both firms and customers are assumed to face prices and they maximize accordingly. The lines marked by M_3 denote the labor market.

If we wished we could introduce one more set of lines representing the payments of dividends $D_1+D_2=\Pi_1+\Pi_2$ back to the consumers. For the purpose at hand this is an unnecessary complication. Furthermore if the production functions are homogeneous of order 1 (i.e., constant returns) then profits will be zero and no dividends will be paid.[2]

Figure 2 shows the two firms (as players 1 and 2) and the two customers (as players 3 and 4). The dotted lines indicate that anyone can deal with anyone directly. There is no imposed market structure or price system.

In particular the two firms and customers can form 2^4-1 or fifteen different trading groups. The trades that these different groups can achieve can be denoted by a characterizing function, (see my forthcoming book with Lloyd Shapley).

$V(\{i\})=0$ is the set of outcomes that a coalition of one individual i can obtain if he deals with nobody else. This amount can be assigned the value 0. It is the status quo payoff. There are four one-person coalitions.

$V(\{i, j\})$ stands for the set of outcomes that a coalition of two individuals can obtain.

$V(\{1, 2\})=V(\{3, 4\})=0$. If the firms or the customers try to trade only among themselves they cannot improve over the status quo.

$V(\{1, 3\})=V(\{1, 4\})=$ the Edgeworth contract curve between the first firm and a single customer (where the customers have the same tastes and endowments).

$V(\{2, 3\})=V(\{2, 4\})=$ the Edgeworth contract curve between the second firm and a single customer.

There are six two-person coalitions possible. Similarly there are four three-person coalitions: $V(\{1, 2, 3\})$, $V(\{1, 2, 4\})$, $V(\{1, 3, 4\})$, and $V(\{2, 3, 4\})$; and there is one four-person coalition whose trading possibilities are characterized by $V(\{1, 2, 3, 4\})$.

II. Orthogonal Coalition Games and Decentralization

Despite its frightening name, the concept of an orthogonal coalition game is relatively straightforward and is the key to the analysis of externalities. The basic idea is that in such a game, once the participants have chosen sides, those excluded from a specific coalition have no influence on the fate of the coalition. This is highly related to the view of open trading in a market. When two traders agree upon a bargain, for all intents and purposes they do not care what the rest of the market is doing. This does *not* mean that the availability of other traders has no influence on the market; however, no one is forced to deal with anyone else. The device of recontracting may be regarded as a comparative statics way of checking trading possibilities; however (as is the case in most actual market places), a contract will be honored by both parties and does not depend upon third parties.

When we view the market with two firms and customers, as described in Section I, as one in which any coalition is feasible then as soon as you have signed your contract with some set of participants the remainder of the market is irrelevant.[3] The essence of an orthogonal coalition game in general and of a market game in particular is "doing your own thing."

---

[2] There is a further complication relating to production. Are the production functions individually owned or can any factory produce anything if it has the inputs? Here we assume the former. This might happen if each owner had a nonmarketable talent or capital good used in the process. A fully satisfactory answer to this question is not attempted here as it is not central to discussion of monetary externalities although it calls for care in treating capacity constraints, nonmarket goods and dividends.

[3] It is possible that, for unspecified reasons, the outsiders might wish to give away extra goods to you,

Consider the coalition $\{1, 3\}$ which can obtain $V(\{1, 3\})$. This is indicated in Figure 2 by V_{13}, and leaves the first firm and customer utterly independent of the second firm and customer. This is not possible in the setup described in Figure 1. The other firm and customer are always interlinked to everyone else through the impersonal market where a price to one is a price to all simultaneously. The market in Figure 1 forces a feedback upon all participants whether they like it or not. They cannot cut out and form their own community. The market in Figure 2 allows for any grouping anyone wants. Once the groups have been decided upon they are strategically isolated. They have no further interaction with each other. In other words they are completely decentralized.

If an organization can be described as an orthogonal coalition game, this means that it is possible to decentralize it in such a complete manner that the performance of any subdivision will depend only upon itself.[4]

Suppose that we were to increase the number of customers and the number of firms of type 1 and type 2, the market we have described is merely a special instance of the market treated by Debreu and Scarf. As the number of firms and customers increases the core shrinks and a price system emerges.

III. Partial Equilibrium Analysis, Comparative Statics and Externalities

The full force of the paradox of pecuniary externalities comes in when we discuss changes in supply costs to an industry considered in isolation; where the changes have been brought about by a shift elsewhere in the economy. For example, suppose that an industry in the same district has had a change in technology which calls for a different type of labor input. The effect of this change may feed back on the labor costs of the first industry.

Following the description of the two models in Section I we can see immediately the effect of the different treatments. Suppose that there was a shift in technology so that the production function for firm 2 becomes $h_2^*(z_2)$ instead of $h_2(z_2)$. Solving the general equilibrium system indicated in Figure 1 a new set of prices will emerge. In particular the cost of the labor supply to Figure 1 will have changed. Viewed from the position of the head of the first firm looking only at his own market, an external force over which he has no control has affected his costs. The reason for this is because we have implicitly assumed that he has to deal via the overall impersonal price system of the mass market whether he likes it or not.

Let us now consider this change in terms of the model in Figure 2. The only values of the characterizing function that have changed are those involving firm 2. If your group was not dealing with firm 2 before its technological breakthrough, it does not have to deal with it now. If it chooses not to, then it will earn the same amount as before.

The effect of the technological change in firm 2's production has not gone unnoticed. It has increased the value of eight out of the fifteen coalitions. By doing so it has made certain coalitions more attractive to some players than they were before and has changed the pattern of worthwhile recontracts. If we were to take the new characterizing function and add more players, the core will now converge to a new limit which is the new price system that takes into account the change in technology.

IV. Markets, Economic Reality, Pecuniary and Physical Externalities

Suppose that exogenous change in the production process of firm 2 were such that it involves a physical externality. For example it switches to a new process which may be labor saving to itself but which dumps smog on the consumers and on the other firm. Returning to Figure 2 and to the characterizing function; we would have to modify the figure by introducing solid interlinkages

---

whether you like it or not. However, if strangers bearing gifts turn up you do not have to take them.

[4] This does not solve the problem of what is the optimal decentralization for the organization as a whole. The answer for a market appears to be: "give up the doing-your-own-thing level of decentralization. Put in a price system in the central office and announce the prices to the otherwise decentralized divisions."

between firm 2 and the three others. Regardless of the market organization they cannot avoid the connection to firm 2. The game is no longer an orthogonal coalition game; we cannot even easily state what a coalition can obtain without first finding out what firm 2 intends to do. It is no longer possible to do your own thing. The characterizing function is not particularly helpful in describing the structure of the threats and interference that the firm with the physical externality can export to all of the other players.

When we try to patch up the characterizing function to take into account the actions of firm 2 we find that the core of the new game may not exist (see Shapley and Shubik 1969). Furthermore if the core were to exist there is no guarantee that it converges as more firms and customers are added. No price system emerges.

In terms of conventional analysis where we *assume* the existence of prices, if we try to solve the general equilibrium system with physical externalities present, although it may be possible to determine prices which clear the markets, they no longer call forth a Pareto optimal outcome.

Returning to Figure 1 we would have to patch up all production functions and the consumer preferences to reflect the role of the smog as an extra commodity. The smog is not an economic good and will not appear with a price, nor can it be traded or avoided. Independent maximizing behavior based upon a price system which does not include a method for pricing and controlling the distribution of the extra commodity will be optimal only by pure coincidence.

In an economy such as ours, markets exist and, at least in some areas, the price system is prevalent. Although the argument in this article has attempted to show that by adopting a more basic approach to trade, based upon the core of an economy and all coalition structures, the phenomenon of pecuniary externalities poses no problem; yet we know that our institutions have markets and we do not deduce their existence *de novo* by calculating the effect of billions of coalitions.

If we assume that firms *must* trade through organized markets, then an analysis of partial equilibrium competition shows that a pecuniary externality may be just as real to a competitor as is smog being dumped on his factory.

If the individual large firm has strategic control over at least part of the environment, its management will not care (unless they are socially motivated) whether costs are going up because a new industry in the area is bidding up the price of engineers or causing costs to rise because it is now more expensive to clean polluted water for the industrial use that the firm has for it. In either case the strategy of interest to the oligopolistic firm is to block the entry of the new industry or to extract a tribute from it to pay for the additional costs it will incur. The firm per se does not care whether (as in the first case) the extra costs are caused by the functioning of a technically efficient price system correctly (from a social point of view) increasing the cost of skilled labor to it; or whether costs are increasing due to a socially undesirable externality which has not been accounted for.

In summary there are three phenomena which must be separated when comparing pecuniary and physical externalities. We may view the whole economy as a mechanistic functioning of the price system. The single price-taking producer will regard a change in costs to him via the price system as bad as a change via a physical externality. However, he is constrained (by being a price taker) to do the right thing when the cost change merely reflects a feedback in the general equilibrium process. He will almost always do the wrong thing when the cost change is caused by a physical externality.

When we view the whole economy as a cooperative game with all coalitions possible, the paradox of the pecuniary externality disappears and is shown to vanish into the overall maximization problem where exogenous change is correctly reflected in the change in the values of some of the outcomes shown in the characterizing function.

When we consider the possibility that an individual firm may be an oligopolist with market power, it is not constrained "to do the right thing" when its costs go up, no matter how they go up. Hence in the sense of an oligopolistic general equilibrium, an individual firm may be expected to use its power to noncooperatively fight any type of perceived externality.

REFERENCES

G. Debreu and H. Scarf, "A Limit Theorem on the Core of an Economy," *Int. Econ. Rev.*, Sept. 1963, *4*, 235–46.

F. Y. Edgeworth, *Mathematical Psychics,* London 1881, pp. 20–25.

L. S. Shapley and M. Shubik, *Competition Welfare and the Theory of Games,* forthcoming, ch. 6.

——— and ———, "On the Core of an Economic System with Externalities," *Amer. Econ. Rev.*, Sept. 1969, *59*, 678–84.

M. Shubik, "Edgeworth Market Games," in A. W. Tucker and R. D. Luce, eds., *Contributions to the Theory of Games,* Princeton 1959.

On the Economics of Mass Demonstrations: A Case Study of the November 1969 March on Washington

By CHARLES J. CICCHETTI, A. MYRICK FREEMAN III, ROBERT H. HAVEMAN, AND JACK L. KNETSCH*

The November 15, 1969 Mobilization to End the Vietnam War in Washington, D.C. represented one of the largest gatherings of people to express dissenting views in the history of the nation. It has been variously estimated that between 119,000 and 1,000,000 people participated in the march and that every state (plus the District of Columbia) was represented. City officials of Washington stated that they were told by the Nixon Administration that the total was 320,000. The Washington police chief estimated the total to be 250,000, "conservatively."[1]

Although these numbers are large, we believe they fail to capture the full significance of the march. Taken alone, they provide no indication of the intensity of feeling among those participating. The fact that many of the demonstrators traveled long distances, involving large time and travel costs, indicates a significant desire to register their feelings in this way. The decision to make such a journey may reveal a greater intensity of feeling than a five-mile drive to attend a local rally or the cost of the stamp and stationery to write one's congressman.

The purpose of this paper is to suggest a better way of assessing this intensity of feeling, or "willingness to pay," in the case of political demonstrations than is possible by examining only the total number of participants. The technique which we employ is the travel-cost method for estimating demand functions.[2] The method has been used to infer from individual choices the willingness to pay for outdoor recreation sites and the economic benefits from public investments in such facilities. The novelty of this paper lies in the application of the technique to participation in a political event rather than consumption of a publicly provided good or service. We suggest that the technique has relevance for models of political choice involving individuals' intensities of feeling on issues as well as their ordinal ranking of outcomes.[3]

We first outline the model, and then describe its application to the November 1969 mobilization based on data gathered on attendance by states. Finally, we discuss the potential and limitations of our model and this approach.

I. The Model

Our procedure is to derive statistically a "demand curve" for participation in the march and to accept the area under this demand curve as a measure of willingness to pay. The travel cost method of estimation is a special case of the general gravity model of social interaction. Other studies have used these types of models to estimate demand functions for transportation, communication, recreation, and tourism.[4] In these models the independent variable V_{ij} is the number of

---

* The authors: are research associate, Resources for the Future, Inc.; associate professor of economics, Bowdoin College; professor of economics, University of Wisconsin; and professor of economics, The George Washington University, respectively. The views presented here are those of the authors and not necessarily of the institutions with which they are affiliated. We wish to acknowledge the clerical assistance of Kathy Crittendon.

[1] The 119,000 estimate was based upon aerial photographs and released by Secretary of Defense Laird. Subsequent to its release, it was withdrawn. The 1,000,000 figure comes from some of the organizers of the march. See the newspaper article by Robert F. Levey.

[2] For an exposition of this method and references to the literature, see Marion Clawson and Jack L. Knetsch, ch. 5.

[3] In such models, opportunities for the formation of coalitions can arise, and vote trading can be an important optimizing mechanism. See Edwin T. Haefele.

[4] For example, see Richard E. Quandt and William J. Baumol; Joseph J. Seneca and Charles J. Cicchetti; and Clawson and Knetsch.

visits, trips, or messages from source i to destination j. The variable V_{ij} is held to be a function of some gravity variable representing the attraction potential for each point, e.g., population N for i and j; other socioeconomic variables S for i and j, and distance D, taken to be a surrogate of the cost of overcoming the separation between source and destination. A general statement of such a model is:

$$V_{ij} = \frac{A N_i^{\beta_1} N_j^{\beta_2} S_i^{\beta_3} S_j^{\beta_4} U_{ij}}{D_{ij}^{\beta_5}} \tag{1}$$

where A is a constant and U_{ij} is a random error term. The modification which we employ hypothesizes that the number of visits from the ith source depends on the attractive force of the destination and the "push" factors of the ith source, say population N_i, socioeconomic variables at source i, and the cost of traveling from i to j. The model can be expressed in terms of the visitation or participation rate and the parameters can be estimated by multiple regression techniques from data detailing the numbers and characteristics of visitors from each source.

Since price is not explicitly stated in the function, a proxy for price must be used to move from the estimated participation equation to the demand curve. In the model as it has been applied, the proxy for price is taken to be changes in travel cost. If a \$1 increase in the price or admission charge for a facility has the same effect on participation as a \$1 increase in travel costs, all the necessary points on the demand curve for the site can easily be calculated. For example, for a price of \$5, the quantity, Q^5, is:

$$Q^5 = \sum_{i=1}^{n} \frac{A N_i S_i^{\beta_1}}{(D_{ij} + 5)^{\beta_2}}, \tag{2}$$

where D_{ij} is price per mile times distance, and the summation is over all visitor sources. Because travel time is a component of travel costs, the value of travel time is included in the empirical estimates presented below. To ignore this component of travel cost would lead to a consistent downward bias in the derived demand function.[5]

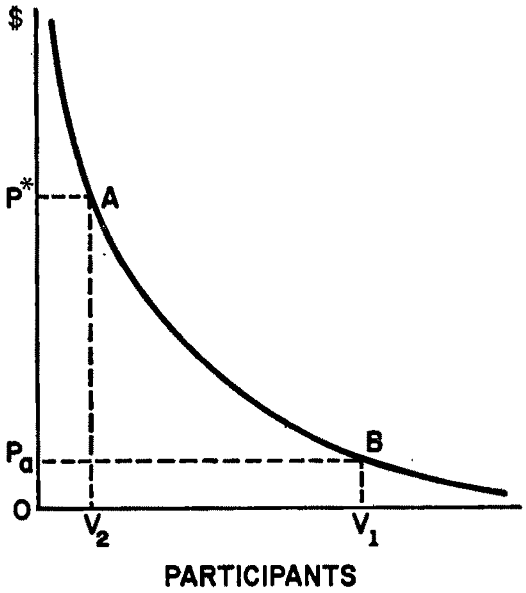


FIGURE 1

Having found the demand curve, we determine the area under the curve by integration. In Figure 1, where P_a is the actual cost of participation (travel time plus mileage, both in dollar terms), the rectangle OP_aBV_1 represents the actual expenditure by all people coming from a particular source. This is the value in exchange or the opportunity costs of the resources (including time) devoted to traveling to the destination and returning. The area under the demand curve but above P_a is an untapped willingness to pay.

II. The Data and the Empirical Results

To estimate the participation equation,

---

[5] Suppose that the money travel cost to a park from City A is \$1 and the trip takes one hour. If a gate fee of \$1 is hypothetically imposed, total money costs become \$2. The conventional model assumes that the people from City A will now have the same participation rate *ceteris paribus* as was recorded for people from City B which is twice as far away, i.e., with travel costs of \$2 and two hours of travel time. But since travel time from City A has not increased, this assumption overestimates the effect of the gate fee in reducing participation. Hence, the estimated demand curve would lie below the true demand curve. To correct for this, it has been suggested that a composite distance variable be formed including both time and money costs, and that only the money cost be hypothetically increased to generate the points on the demand curve. See Frank J. Cesario and Knetsch.

data on participants, population, travel cost, travel time, and other socioeconomic variables are required for each source. Each of the thirty-two states east of and bordering on the Mississippi River (plus Texas and the District of Columbia) was considered a separate source. Data on the number of participants from each source came from a questionnaire mailed to the student body president, the editor of the student newspaper, and the chairman of the sociology department of each of a selected sample of over 200 colleges and universities. The questionnaire asked for an estimate of the total number of participants from that state; the response rate was about 45 percent. Because estimates from some states varied substantially, the median of all estimates was chosen for each state. Population is for the year 1967. Distance and travel time came from the Rand McNally Road Atlas.[6] Travel costs are assumed to be 5.5¢ per person per mile based on operating costs for private automobiles.[7] The opportunity cost of travel time is taken to be $1 per hour.[8] To reflect differences in the composition of the populations and differences in political attitudes among states, college student enrollment in the state as a percentage of population and the combined Nixon-Wallace 1968 presidential vote in the state as a percent of the total state vote were used in some of the regression estimates.

The participation relationship was estimated in several functional forms with alternative specifications of the exogenous variables. All of the equations are double logarithmic regression equations (see Table 1). An instrumental variable technique analogous to two-stage least squares was used for estimation to avoid the asymptotic bias and inconsistency of ordinary least squares.[9]

In each of the equations the percentage of students in each state is positively related to the state participation rate. In equations C and D the coefficient for the state Nixon-Wallace vote percentage is negative, indicating an inverse relationship between the moderate and conservative political position and participation in the march. Equations A and C have the additive $(M+T)$ measure for the travel (distance and time) variable, while equations B and D use the multiplicative form $(M \cdot T)$.

All four equations are statistically significant, and all account for about 70 percent of the variation in the dependent variable. Because the additive form for the travel cost variable generates a lower demand function than the multiplicative form, it was chosen to estimate the demand function.[10] Moreover, although both equations A and C predict the number of marchers to be 316,900, equation A explains a higher percentage of the variation in the total *number* of marchers from each state than does equation C.

Since the regression was estimated in the logs, the demand curve is asymptotic to the vertical axis. Thus, consumer surplus is undefined. To obtain an estimate of consumer surplus, the demand curve above P_a for each state was integrated from the actual number of state participants (OV_1 in Figure 1) to 5 percent of the actual number of participants (OV_2). Then the rectangular area OP^*AV_2 was added for the percent of the participants not included in OV_2-OV_1. This procedure, therefore, excludes some portion of the full consumer surplus of the 5 percent of the participants of each state with the

---

[6] Distance and travel time was from the largest city in the state, except for Florida, Tennessee, and North Carolina. For these states, geography and population distributions suggested that Tampa, Nashville, and Raleigh would be more appropriate points from which to measure distance. Also, the population of the adjoining Maryland and Virginia counties was allocated to the District of Columbia.

[7] The data on operating costs of automobiles is from *Cost of Operating an Automobile.* On the basis of supplemental data on the per person per mile rates of air, rail and charter bus travel, the 5.5¢ figure appears justified as a "best guess."

[8] Participation and travel cost data by state are presented in the Appendix.

[9] Since both the dependent and some independent variables are deflated by population, ordinary least squares estimation would violate one of the classical least square assumptions [$E(X'U)$ must equal 0] and generate "least squares bias." The instrumental variable technique adjusts for this effect and ensures asymptotic unbiasedness.

[10] Because the multiplicative form implies a convex trade off function between time and travel costs, it yields higher estimates of participation for each increase in assumed cost than does the additive form.

TABLE 1—EMPIRICAL RESULTS OF DEMAND ESTIMATION MODEL*

Dependent Variable: participants as a percentage of state population	Constant	Students as a percent of state population	Combined percentage of state votes for Nixon and Wallace	Travel cost variable (additive)	Travel cost variable (multiplicative)	R^2/F
Equation A	−12.3493	4.1136	—	−1.1459	—	.68
	(− 2.4)	(2.8)	—	(−7.0)	—	34.4
B	−12.3797	4.1911	—	—	− .5255	.68
	(− 2.4)	(2.9)	—	—	(−7.1)	35.5
C	−10.4355	2.9344	−1.9303	− .9034	—	.70
	(− 2.1)	(1.9)	(−1.9)	(−4.4)	—	35.5
D	−11.2117	3.3809	−1.3429**	—	− .4384	.69
	(− 2.2)	(2.1)	(−1.2)	—	(−4.2)	24.3

(Note: *t*-values appear in parenthesis below their respective coefficients.)
* All variables in natural logs.
** Marginally significant at about the 80 percent level.

highest willingness to pay. The value in exchange was determined by multiplying the number of participants (OV_1) predicted by regression equation B by the total travel cost (OP_a) including imputed travel time costs and equals $OP_a \cdot OV_1$.

The total consumer surplus of the event was estimated by subtracting value in exchange of each state from the total area under the demand curve of each state $(OP^*ABV_1 - OP_aBV_1)$ and aggregating over all of the states. This procedure yields an estimate of $10.1 million. Based on the 5.5¢ per person per mile and $1 per hour opportunity cost, the value in exchange is $7.1 million.[11] The total willingness to pay for participation in the event—the sum of consumer surplus and value in exchange—is $17.2 million, which is slightly more than $54 per marcher.

III. Conclusions

Although we have estimated a participation function, used it to generate a demand curve, and calculated the area under it, the question of interpretation remains. There are several possible interpretations. The logic of our technique tells us that we have estimated the aggregate willingness of the participants in the march to pay for the opportunity to express their views in the context of that event. This statement, however, does not distinguish between the event per se and the sentiments which moved people to participate. Our data provide no empirical basis for making such a distinction. It should be emphasized that our estimates imply nothing regarding the benefits and costs experienced by nonparticipants because of the occurrence of the march. In the case of nonparticipating beneficiaries, the "free-rider" phenomenon may indicate the potentiality of substantial benefits.

At one extreme the march could be viewed as a consumer good and participation in it the consequence of utility maximizing decisions no different from the decision to purchase a bottle of wine. Conventional welfare analysis would then lead us to conclude that the net benefit of the march was the consumer surplus ($10.1 million) less any external costs.[12] This position, however, would likely misrepresent the significance of the event and the motivations of the people who came to Washington for it. While many of the participants felt that their participation was enjoyable, if not rewarding and at times moving, it is unlikely that many would have come if all there was to it was the prospect of a walk, a few hours in the autumn chill, and a whiff of tear gas to cap the day.

---

[11] While the coefficients of the explanatory variables are insensitive to time and travel cost values, the constant term of the estimated equation, and therefore the estimate of benefits, is sensitive to the values chosen. For example, if travel costs are taken to be 4¢ per person per mile, the estimate of the area under the demand curve is reduced from $17.2 million to $13.5 million. We feel that an estimate of value in exchange of from $5 to $10 million and of consumer surplus of from $10 to $20 million is a valid interpretation of the data we have analyzed in this study.

[12] The costs borne by others because of the November march have been estimated to be $1.8 million by Representative Gerald R. Ford of Michigan. This figure includes: $240,000 for private buildings that were damaged; $28,000 for government property damage, $473,776 for additional overtime for law en-

At the other extreme, participation in the event could be viewed purely as an act of political expression. Then the area under the demand curve (including value in exchange) would represent an intensity of political feeling or a willingness to commit resources to achieve a desired political end. In this light, participation involved the commitment of resources to an act of production with the desired output being a specific change in public policy.[13]

There are other possible interpretations. For example, one could view this measure as an indication of untapped political capital that could be mined by candidates or organizations advocating particular policies. In this perspective, the $8 million debt of the National Democratic Party in 1970 is eclipsed by our $17.2 million estimate of willingness to pay.

To summarize, we have presented an estimate of the revealed willingness to pay for the opportunity of participating in a particular political demonstration and have suggested that this estimate is an indication of the intensity of feeling of participants toward the political issue. Our results illustrate a technique for evaluating the feelings of those people acting in a particular way on a particular issue and clearly not the feelings of all people toward this issue. However, *ex post* analyses of the behavior of individuals toward political issues do have meaning as indicators of the intensity of political feeling, especially in the context of models of vote trading, coalitions, and advocacy.

APPENDIX

State	Median Number of Marchers	Average Round Trip Cost[a] (Time and Distance)
Alabama	200	$118.28
Arkansas	85	166.17
Connecticut	15,000	52.62
Delaware	1,500	15.11
Florida	1,700	147.78
Georgia	250	96.46
Illinois	6,500	109.57
Indiana	1,000	87.54
Iowa	325	159.48
Kentucky	375	99.90
Louisiana	75	173.89
Maine	500	83.14
Maryland[b]	12,500	5.95
Massachusetts	15,000	68.07
Michigan	15,000	74.87
Minnesota	1,000	173.01
Mississippi	25	157.56
Missouri	725	135.95
New Hampshire	500	73.92
New Jersey	5,000	32.76
New York	50,000	35.09
North Carolina	14,000	34.28
Ohio	5,000	53.61
Pennsylvania	26,250	20.96
Rhode Island	1,500	62.33
South Carolina	340	73.43
Tennessee	950	110.51
Texas	1,000	223.26
Vermont	1,000	83.34
Virginia[b]	6,000	16.49
West Virginia	700	59.22
Wisconsin	6,750	122.86
Metropolitan D.C.	103,000	1.12
	300,000	

[a] This estimate is based on the use of 5.5¢ per person per mile as the travel costs and $1 per person per hour for the time costs implied by the travel.

[b] Not suburban District of Columbia.

---

forcement and clean-up personnel, and $963,088 to maintain troops and National Guardsmen in the city. No doubt there were other positive and negative consumption externalities. One respondent said, "Thanks for having it in Washington. Our campus was clean and quiet that weekend."

[13] Some credence is given to this interpretation by both the feeling that after November 15th the mass demonstration as a tactic was dead and the observed outpouring of energy and time into more conventional modes of political activity after President Nixon's announcement of the Cambodian invasion.

REFERENCES

F. J. Cesario and J. L. Knetsch, "The Time Bias in Recreation Benefit Estimates," *Water Resources Research,* June 1970, *6,* 700–04.

M. Clawson and J. L. Knetsch, *Economics of Outdoor Recreation,* Baltimore 1966.

E. T. Haefele, "A Utility Theory of Representative Government," *Amer. Econ. Rev.,* June 1971, *61,* 350–67.

R. F. Levey, *Washington Post,* Dec. 20, 1969.

R. E. Quandt and W. J. Baumol, "The Demand for Abstract Transport Modes: Theory and Measurement," *J. Reg. Sci.,* 1966, *6,* no. 2, 13–26.

J. J. Seneca and C. J. Cicchetti, "A Gravity Model Analysis of the Demand for Public Communication," *J. Reg. Sci.,* 1969, *9,* no. 3, 459–70.

U.S. Department of Transportation, Federal Highway Administration, Bureau of Public Roads, *Cost of Operating an Automobile,* Feb. 1970, Washington.

The Graduated Fair Return: Comment

By C. ROBERT WICHERS*

Harvey Averch and Leland L. Johnson assert that rate-of-return regulation stimulates excessive use of capital so that the regulated firm produces inefficiently. To keep capital expansion down, Alvin K. Klevorick has suggested making the permitted rate of return a decreasing function of capital. That way the firm cannot increase its use of capital without bringing down the regulatory ceiling and thus punishing itself. The purpose of this note is to show that Klevorick's analysis contains a flaw which makes the graduated fair return proposal inoperative.

The main assumptions of the Averch-Johnson model are: a firm maximizes profit, operates under completely stationary conditions, employs two inputs K (capital) and L (labor) for which it pays fixed prices. Furthermore, the firm produces one product and sells it at a price which depends on output; output is zero when either K or L is zero.

Let R, r, and w represent revenue, cost of capital and the cost of labor, respectively. Profit is then $R-rK-wL$, and the rate of return on capital is $(R-wL)/K$. The regulatory agency determines the greatest value, s_0, which this magnitude may have: $R-wL \leq s_0K$. Profit is maximized subject to this constraint.[1] For the sake of simplicity, we assume that a maximum exists and is unique and that s_0 is neither so small as to make profits vanish (that is, we assume $s_0>r$) nor so large as to render the constraint inactive.

Let R_K, R_L abbreviate $\partial R/\partial K$, $\partial R/\partial L$. We differentiate the Lagrangean expression

$$(1)\quad (R - rK - wL) - \lambda(R - s_0K - wL)$$

and set the derivatives equal to zero:

$$(2)\qquad (R_K - r) - \lambda(R_K - s_0) = 0$$

$$(3)\qquad (R_L - w) - \lambda(R_L - w) = 0$$

$$(4)\qquad R - s_0K - wL = 0$$

If (K_0, L_0, λ_0) is the solution to this system, then K_0, L_0 are the inputs which maximize profit (subject to the assumedly active constraint). It can be shown that $0<\lambda_0<1$, and Averch and Johnson use this to prove that R_K is less than r. Since $R_L=w$, by (3), the firm operates inefficiently: R_K/R_L does not equal r/w. In fact, it is less than r/w. Averch and Johnson also assert that capital input expands when regulatory control is imposed. This was later proved by Akira Takayama who through differentiation of (4) showed that dK_0/ds_0 is negative.

Klevorick proposes making the permitted rate of return a function of capital, $s=s(K)$ instead of $s=s_0$. Moreover, $s'=ds/dK$ must be negative at K_*, the amount of capital employed under this regulatory scheme. Intuitively one expects K_* to be less than K: the firm will be less eager to expand its capital when by doing so it turns on the regulatory screws. On the other hand, Takayama's proof that a lower value of s_0 induces the use of more capital seems to suggest, intuitively again, that the opposite will occur. In that case, the process would feed on itself and the amount of capital employed would soon increase beyond all bounds. The mathematics will have to settle whether Klevorick's proposal mitigates or reinforces capital expansion (or does neither).

When s_0 is replaced by $s(K)$, equations (2), (3), and (4) become

$$(2')\quad (R_K - r) - \lambda(R_K - s - Ks') = 0$$

$$(3')\qquad (R_L - w) - \lambda(R_L - w) = 0$$

$$(4')\qquad R - sK - wL = 0$$

We assume that this system has exactly one solution (K_*, L_*, λ_*) and that $0<\lambda_*<1$. In fact, for the purpose of this note it is sufficient to have $\lambda_* \neq 1$, so that assuming the existence of exactly one solution is enough, see (3′).[2] Clearly, the values of K_*, L_*, and λ_* depend on the shape of $s(K)$, and straightforward comparison of K_*, L_* with K_0, L_0 is not necessarily meaningful. To make it meaningful Klevorick proposes choosing $s(K)$ so that

---

* Assistant professor of economics, State University of New York at Stony Brook.

[1] Several of these assumptions can be relaxed without affecting the conclusion drawn by Averch and Johnson. For instance, the number of products may be arbitrarily large; so may be the number of inputs. Similarly, depreciation (which the form of the constraint shows to be zero) may have a positive value.

$$s(K_*) = s_0 \tag{5}$$

and

$$\lambda_* = \lambda_0 \tag{6}$$

What do these conditions imply?

Dividing both sides of (3) and (3′) by the assumedly nonzero number $(1-\lambda)$ shows that

$$R_L = w \tag{7}$$

at (K_0, L_0) as well as at (K_*, L_*). Apparently, K_0 and L_0 are determined by (4) and (7) while K_* and L_* are determined by (4′) and (7). Setting $s(K_*)=s_0$ to satisfy the first condition makes the distinction between (4) and (4′) disappear. It follows that $K_0=K_*$ and $L_0=L_*$: regulation according to Klevorick's proposal has the same effect as regulation with a constant rate-of-return ceiling. Furthermore, this conclusion does not involve λ_*, suggesting possible inconsistency of conditions (5) and (6). Indeed, satisfaction of both $s(K_*)=s_0$ and $\lambda_*=\lambda_0$ implies that $s'(K_*)=0$, as can be seen from (2′); but s' is assumed to be negative at $K=K_*$.

REFERENCES

H. Averch and L. L. Johnson, "Behavior of the Firm under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

A. K. Klevorick, "The Graduated Fair Return: A Regulatory Proposal," *Amer. Econ. Rev.*, June 1966, *56*, 477–84.

A. Takayama, "Behavior of the Firm Under Regulatory Constraint," *Amer. Econ. Rev.*, June 1969, *59*, 255–60.

---

[2] If $\lambda_*=1$, the amount K_* of capital is determined by the shape of the function $s(K)$ chosen by the regulator. Therefore, Klevorick argues, $\lambda_* \neq 1$ must be assumed if we are intent upon examining the behavior of the *firm*. But this is insufficient reason: K_* is determined not solely by the shape of $s(K)$ but also by the supposed profit maximization.

The Graduated Fair Return: A Further Comment

By Alvin K. Klevorick*

C. Robert Wichers has brought to light an important shortcoming in my original analysis of the graduated fair return proposal. I am indebted to him for his careful reading of my paper and for making the discovery he has. The essence of my proposal was that regulation could be improved if the fair rate of return s were converted from a constant s_0 into a decreasing function of capital $s=s(K)$, with $s'(K)<0$. In attempting to prove that this modification of regulatory procedures would decrease the input inefficiency noted by Harvey Averch and Leland L. Johnson, namely, the divergence between the marginal rate of substitution (*MRS*) of capital for labor *and* the factor-cost ratio, the comparison I drew between the constant and graduated fair return methods of regulation was invalid.

This comparison[1] required that: (i) the value of the graduated fair return at the firm's optimal position (K_*, L_*) be equal to the value of the constant fair rate of return, that is, $s(K_*)=s_0$; and (ii) the shadow price for the constant fair rate constraint be equal to the shadow price for the graduated return constraint, that is $\lambda_*=\lambda_0$. As Wichers argues, the first requirement can be met only if the graduated fair return schedule is constructed so that s_0 equals $s(K_0)$, where K_0 is the optimal capital stock in the constant fair return model. But, then, the optimal K, L pair for the graduated fair return problem will be identical with the optimal pair for the constant fair return problem, and a graduated fair return causes no change in input efficiency. Imposing requirement (ii) above engenders a further inconsistency. For, as Wichers notes, one cannot simultaneously require $s(K_*)=s_0$, $\lambda_*=\lambda_0$ *and* have any degree of graduation at the optimum (that is, have $s'(K_*)\neq 0$).

In light of Wichers' arguments and after further consideration, I no longer believe that a persuasive case can be made for either of the two conditions my 1966 paper asserted were necessary for a meaningful comparison of the alternative regulatory schemes. When one rejects these conditions for a comparison, there remains, however, an open question. Namely, suppose a fair rate of return regulatory constraint is in effect with the fair rate at the constant level s_0. Assume, then, that this constraint is replaced with one in which the fair rate varies with the amount of capital input. What effect, if any, would this change have on the gap between the *MRS and* the ratio of factor costs?

To answer this question, recall that with a fair rate of return which depends on the capital used, the firm's optimal position is, following Wichers' notation, described by (1)–(3):

$$(1)\quad (R_K - r) - \lambda(R_K - s(K) - Ks'(K)) = 0$$

$$(2)\quad (R_L - w) - \lambda(R_L - w) = 0$$

$$(3)\quad R - wL - s(K)K = 0$$

For the case where the constraint (3) is effective and the marginal return to capital always exceeds the market cost of capital, one obtains

$$(4)\quad 0 < \lambda < 1, \qquad \frac{R_K}{R_L} < \frac{r}{w}, \qquad R_K < r < s(K) + Ks'(K)$$

Measuring input inefficiency by the divergence between R_K/R_L (which equals $\partial L/\partial K$, the *MRS* of capital for labor) and r/w, it is clear that since r and w are constant, to reduce this inefficiency it is necessary to increase R_K/R_L. From (2), though, it follows that

$$(2')\quad R_L = w$$

no matter how $s(K)$ is set *so long as* $\lambda\neq 1$. Thus, *to reduce the gap between the marginal*

---

* Associate professor of economics, Yale University.

[1] The notation adopted throughout this paper is that of Wichers.

rate of substitution and the factor-cost ratio an increase in R_K is required.

Taking the simplest case, suppose the rate of change of the fair return depends on only one parameter, α. When $\alpha = 0$, there is no graduation of the rate of return; and as α increases, the fair rate and the speed of graduation at a *given* capital input increases. Let the fair rate of return for any level of capital input be given by the function

$$s(K) = \bar{s} + f(K, \alpha) \tag{5}$$

$$\text{with} \begin{cases} f_\alpha = \dfrac{\partial f}{\partial \alpha} > 0, & f(K, 0) = 0 \\ f(0, \alpha) = 0, & f_{K\alpha} = \dfrac{\partial^2 f}{\partial K \partial \alpha} > 0 \end{cases}$$

The conditions on f_α, $f_{K\alpha}$, and $f(K, 0)$ formalize the definition of α, while the fact that $f(0, \alpha) = 0$ means that the fair rate of return on the first unit of capital is $\bar{s}$. If there is no graduation, i.e., $\alpha = 0$, $s(K)$ is constant at $\bar{s}$. If $f_K > 0$, then the fair rate is graduated upward so that each additional unit of capital brings a higher average rate of return for capital, while if $f_K < 0$, the fair rate is graduated downward, as I originally argued it should be.

To investigate what proper construction of this $s(K)$ function can accomplish, note first that (2′) and (3) form a subsystem determining the optimal values of K and L.[2] Hence, to determine the (differential) changes in K and L that result from differential changes in $\bar{s}$ and α, the policy parameters of this $s(K)$ function, use equation (5) to differentiate (2′) and (3) totally with respect to $\bar{s}$ and α. The differential change in R_K then follows simply from the fact that $dR_K = R_{KK}dK + R_{KL}dL$, where $R_{KK} = \partial^2 R/\partial K^2$ and $R_{KL} = \partial^2 R/\partial K \partial L$. Proceeding through the algebra, one obtains

$$dR_K = \left(\frac{R_{KK}R_{LL} - R_{KL}^2}{R_{LL}}\right) \cdot \frac{K(d\bar{s} + f_\alpha d\alpha)}{(R_K - \bar{s} - f - f_K K)} \tag{6}$$

Assuming $R(K, L)$ is concave in K and L, the first fraction is negative. From (4), $R_K - s(K) - s'(K)K < 0$, and for $s(K)$ as defined in (5) this means $R_K - \bar{s} - f - f_K K < 0$. Combining these two pieces of information, we have:

$$\operatorname{sign} d\left(\frac{R_K}{R_L} - \frac{r}{w}\right) = \operatorname{sign} dR_K = \operatorname{sign}\,(d\bar{s} + f_\alpha d\alpha) \tag{7}$$

Any combination of changes in the fair rate of return schedule for which $d\bar{s} + f_\alpha d\alpha$ is positive will reduce the (algebraic) gap between the *MRS* and the factor-cost ratio. For example, beginning with a constant fair rate of return $\bar{s}$, to reduce input inefficiency as measured by $R_K/R_L - r/w$, one could:[3] a) leave $\bar{s}$ unchanged and graduate *upward*; or b) with proper choices of $d\bar{s}$ and $d\alpha$, decrease $\bar{s}$ and graduate upward even more rapidly; or c) increase $\bar{s}$ and graduate the fair return downward.

But equation (7) also indicates a much simpler way of reducing input inefficiency as measured by $R_K/R_L - r/w$. If one begins with a constant fair rate of return $\bar{s}$, a reduction in the gap between the *MRS* and the factor-cost ratio can be achieved by setting $d\alpha = 0$ and $d\bar{s} > 0$: simply increase the level of the constant fair rate of return. This observation makes clear exactly what graduation of the fair rate of return can and cannot accomplish. Specifically, graduation of the fair rate of return is a perfect substitute—nothing more, nothing less—for shifting the level of a constant fair rate in order to reduce input inefficiency as measured by the gap between the marginal rate of substitution between factors *and* the factor-cost ratio.

While it is to be hoped that Wichers' paper and the present one have helped to clarify the issues surrounding graduated fair return regulation, one important issue has gone unmentioned thus far. The entire discussion has focused on the effect of regulation on the input quantities chosen by the regulated firm. In this, the present discussion has followed the pattern of much of the theoretical literature which developed from the original Averch-Johnson paper. With few exceptions, notably, Alfred E. Kahn's comment on the graduated fair return proposal, these papers elevate input efficiency from the status of a means to an end to the status of an end in itself. Any meaningful attempt to select a level at which a constant fair rate of return, or a schedule of fair rates, should be set must, however, recognize that input efficiency is but a means to an end and must confront the trade off made explicit by Kahn.

---

[2] Equation (1) then determines the optimal value of λ. Wichers uses this same observation with regard to subsystem (4) and (7) *and* subsystem (4′) and (7) in his paper.

[3] Recall that $f_\alpha = \partial f/\partial \alpha$ is positive by definition (5).

Specifically, suppose with Kahn that as the fair rate of return decreases, the output of the regulated firm increases.[4] Then reductions in the gap between the marginal rate of substitution and the factor-cost ratio that are achieved by raising the fair rate of return, or by substituting a properly graduated fair return for a constant one, will only come at the cost of a reduction in output. The social welfare *gain* resulting from improved input efficiency must be weighed against the social welfare *loss* resulting from the output restriction. Only an analysis which takes explicit account of both of these effects can hope to help regulators in choosing the correct level (or schedule) for the fair rate of return.[5]

REFERENCES

H. Averch and L. L. Johnson, "Behavior of the Firm under Regulatory Constraint," *Amer. Econ. Rev.*, Dec. 1962, *52*, 1052–69.

A. E. Kahn, "The Graduated Fair Return: Comment," *Amer. Econ. Rev.*, Mar. 1968, *58*, 170–73.

A. K. Klevorick, "The Graduated Fair Return: A Regulatory Proposal," *Amer. Econ. Rev.*, June 1966, *56*, 477–84.

———, "The 'Optimal' Fair Rate of Return," *Bell J. Econ.*, spring 1971, *2*, 122–53.

E. Sheshinski, "Welfare Aspects of a Regulatory Constraint: Note," *Amer. Econ. Rev.*, Mar. 1971, *61*, 175–78.

---

[4] This will be the case so long as capital is not a regressive input.

[5] I have attempted to pursue such an investigation in my *Bell Journal* article. Eytan Sheshinski takes a similar approach to the problem of optimal regulatory control.

Mass Balance, General Equilibrium, and Environmental Externalities

By ROGER G. NOLL AND JOHN TRIJONIS*

Recently Robert V. Ayres and Allen V. Kneese developed a model for examining the externalities associated with the deposition of residuals from production and consumption processes. The main point of their paper, that a general equilibrium analysis is required if the secondary effects of pollution and pollution abatement strategies are to be given proper consideration, is incontestable (although in certain specific instances the importance of secondary effects may not be great enough to be worth examining). Nevertheless, the Ayres-Kneese model set forth to demonstrate this proposition has several shortcomings, some of which lead the authors to neglect certain important environmental problems and avenues for abating pollution, and others of which lead to erroneous conclusions about the nature of the pollution problem.

Ayres and Kneese present two related analytical devices for analyzing the production of pollutants and the diseconomies of pollution: the "mass balance" concept and the Walras-Cassel general equilibrium model, both of which assume linear relationships among inputs and outputs. Ayres and Kneese recognize the limitations of their linearity assumption about production functions; by disallowing factor substitution, they afford no opportunity for producing a given commodity by using fewer natural resources and more services of men and machines.[1] They are also aware of one limitation of static equilibrium analysis; no possibility exists for examining mechanisms to induce improvements in the efficiency of production processes through technological changes which reduce natural resource requirements (such as by improving the degree of completeness of combustion in an automobile engine), although new techniques, once known, can be added to the set of permissible production functions incorporated into the static equilibrium model. But the limitations of the Ayres-Kneese approach go far beyond the implications with respect to production. In particular, the assumption that inputs, outputs, pollution, and consumer satisfaction bear a consistent relation to physical mass is, in several important instances, misleading. Our primary purpose is to show several modifications to the Ayres-Kneese model that make the model more representative of the real problems associated with pollution and more useful for planning pollution abatement.

I. Mass Balance

The mass balance concept as applied to pollution originated in the analysis of particular disposal mediums, especially water, and in the study of particular pollutants. In water system analysis, mass balance refers to the simple concept that the mass of salts, nutrients, pollutants, etc., added to a water system must equal the mass which leaves

---

* Roger G. Noll is Senior Fellow, Brookings Institution and associate professor, California Institute of Technology (on leave). John Trijonis is a graduate student in environmental engineering science, California Institute of Technology. We are grateful for useful comments on an earlier draft by Henry Aaron, Ralph d'Arge, Burton Klein, and Allen Kneese. The costs of preparing the manuscript were financed by the Ford Foundation through Brookings Studies in the Regulation of Economic Activity. The views expressed in this paper are not intended to represent the views of the trustees, officers, or staff of the Brookings Institution.

[1] Ayres and Kneese do miss one very important consequence of this assumption; the impossibility of using their model to test the value of flexible technologies for dealing with pollution through time. Following George Stigler's ideas about "adaptability" in production processes, some built-in flexibility as a contingency against future environmental problems might be worth the cost. For example, automobile engines capable of using several different kinds of fuel would, at some cost, permit policy to be readily adapted to changing characteristics of the air pollution problem.

(at a purification plant, by deposition in the ocean, etc.). In the analysis of a particular pollutant, mass balance analysis is the process of tracing a given mass of that pollutant as it proceeds through the environment. For instance, lead pollution may be analyzed by identifying the sources of lead and then following the mass emitted by these sources as it is transported through the environment. Again, the mass of lead remains the same throughout.

Ayres and Kneese extend the mass balance concept to the analysis of all pollutants and relate pollution to the inputs of mass into the production process. Let M_I be the mass flow from nature of inputs to production and consumption processes, M_R be the mass flow of residuals from production and consumption back to the natural environment, M_P be the mass of pollutants creating external diseconomies, and M_A be the mass accumulated through net capital formation. Since mass must be conserved, $M_I = M_R + M_A$.[2] The crux of the pollution externalities problem is how the residuals relate to pollutants. Obviously, the mass of residuals is not equal to the mass of pollutants since most residuals (diatomic oxygen, for example), have no deleterious environmental effect. Ayres and Kneese assume that nature has a given assimilative ability for wastes, M_0, and that pollution is the residual mass that exceeds it. Further, they assume that the excess of residual mass above M_0 is the basic measure of the level of pollution:

> . . . technological means for processing or purifying one or another type of waste discharge do not destroy the residuals but only alter their form. Thus, given the level, patterns, and technology of production and consumption, recycle of materials into productive uses or discharge into an alternative medium are the only general options for protecting a particular environmental medium such as water. [p. 283]

Two refinements of these assumptions should be made. First, M_P, M_R, and M_0 should be vector-valued quantities, the components of which are specific residuals and pollutants. Through the vector approach one can account for the fact that nature has varying assimilative capacities for different residuals. Some residuals, such as O_2 or H_2O given off in production processes, have very large natural assimilative capacities or very low disutility, and therefore cause no significant pollution. Contrary to the quotation above, a third alternative to protecting a given environmental medium is to change the *form*, but not necessarily the *mass*, of residuals.

Second, if one uses the vector approach, mass is not the most useful measure of the vector components. In most instances the number of molecules of a pollutant is the most relevant measure of pollution, since the damage caused by pollution is usually through a chemical reaction. For example, the corrosive power of oxidants is not generally related to the atomic weight of the molecules. Thus, knowing that the atmosphere contains X parts per million strong oxidants is more useful than knowing it contains Y tons, especially if one is not told the specific proportions of Y due to each oxidizing pollutant. An even more useful measure than parts per million would include the reactivity of molecules. This suggests a useful, intermediate degree of aggregation between the vector of specific residuals and the mass of pollutants which would be a vector of pollution intensity by types of pollutant. For example, the amount of each oxidant could be weighted by its reactivity to form a weighted sum for measuring the level of "oxidant pollution."

Still another factor must be considered if residual measurements are to be connected to measures of pollutants. The elements of the vector of residuals often interact to form a new pollutant or to alter natural residual absorption capacity. For instance, assimilative capacity for one residual will in general depend on the mass flow of other residuals. Furthermore, the damage due to a particular pollutant will depend on the mass flow of other pollutants. These concepts are illustrated in the Los Angeles smog problem. The pollutants in photochemical smog are interdependent in that they react chemically with one another. Los Angeles may have a high assimilative capacity for hydrocarbon emissions as long as emissions of oxides of nitrogen are small; however, an increase in oxides of nitrogen brings about chemical reactions which produce irritating aldehydes and oxidants. Thus, both the tolerance level for hydrocarbons and the damage associated with their emission are affected by the level of nitrogen oxides, and vice versa. One way of incorporating this phenomenon into a general equilibrium model is to separate residuals from pollutants by adding a new set of equations that represent transformation of the residual vector into a vector of pollutants. Some of these equations would be chemical formulas, others would be identities representing residuals that are also pollutants. Some of the independent variables in these equations would be natural conditions which affect the type of pollutant produced by a residual.

---

[2] Ayres and Kneese neglect elemental oxygen deposition in the atmosphere, which in their model would make $M_I > M_R + M_A$, although they explicitly include elemental oxygen as an input. To be strictly correct, O_2 given off in production processes is as much a residual as its pollutant brother, O_3 (ozone). Ayres and Kneese also neglect a number of environmental problems unrelated to mass flow, such as thermal pollution, radioactivity (nuclear reactions do not conserve mass), and visual ugliness or aesthetic pollution.

Because of its increased generality, the above analysis admits a wider scope in the search for methods of abating pollution than decreasing the total mass of residuals through recycling or increasing total assimilative capacity through public investment. The recognition of the interdependencies among residuals opens up the alternative of exploiting these relations to find more effective methods of dealing with pollution. One general principle following from this is that society should put effort into controlling pollutants that tend to decrease the assimilative capacity and increase the disutility of other residuals. These new alternatives will not necessarily reduce the mass of residuals, and so they are possible pollution abatement strategies that are excluded by an approach that assumes mass to be the relevant measure of pollution.

II. General Equilibrium

The second theoretical tool developed by Ayres and Kneese is an extension of the Walras-Cassel model to keep track of the mass flow of all materials in a manner consistent with the concept of mass balance. The end product of their model is the set, $\{Y\}$, of final demands, measured in units of mass. The mass of the final demands is assumed to be the measure of consumer services. That is, consumer services are implicitly taken to be determined by the mass of consumer goods through a fixed functional relationship; the arguments of individual utility functions are masses of goods. Technology being fixed, the set $\{Y\}$ is related to mass inputs through production functions of the form:

$$Y_i = f_i(r_1, \dots, r_j, \dots, r_L, \quad s_1, \dots, s_k, \dots, s_P) \quad i = 1, \dots, N \tag{1}$$

where r_j equals mass-flow input of the jth resource ($\sum_j r_j = M_I$) and s_k equals the services from the kth productive factor. If C is consumer satisfaction, and if there is no possibility of substitution among the r's and the s's, then $C = C(r_j)$, a fixed functional relation —in the Walras-Cassel model, a fixed coefficient, linear relation—between consumer satisfaction and the mass of resources used.

Once one assumes a fixed relation between the level of pollution and residual mass and another fixed relation between consumer services and mass inputs, one has a fixed relation between consumer services and pollution. In this situation pollution can be reduced while maintaining full employment of resources in only three ways. First, ". . . the assimilative capacity of environmental media can sometimes be altered" (p. 295), i.e., M_0 can be increased. Thus Ayres and Kneese recommend ". . . the planning and execution of investments with public goods aspects" (p. 295), such as river flow augmentation. Second, the mix of total output can be altered by shifting consumer demand to goods and services requiring less mass. Presumably this would occur if environmental damage shadow prices were actually charged.[3] Third, pollution can be abated by recycling the residual mass back into the economy. In this way, the actual residual mass deposited in the environment can be reduced and M_P decreased.

---

[3] As A. O. Converse observes, charging for environmental damage would also lead to the use of less polluting production technologies.

The connection between consumer services and mass flow assumed by Ayres and Kneese not only implies a fixed "consumption technology," but also implies a one-to-one correspondence between consumer goods and the goods characteristics that are the arguments of individual utility functions and hence the basis for consumer demand. The assumption eliminates from consideration the promising alternative to be inferred from Kelvin Lancaster's important contribution to the theory of consumer choice that the relationship between mass flow and consumer services is not fixed. If the mix of goods providing a given characteristic is allowed to change, it becomes possible to decrease the mass flow for a given level of consumer satisfaction. For example, public transportation systems may be able to provide the same service for a lower mass of inputs compared with private automobiles. Or newspapers may be broadcast over cable television systems instead of printed on paper. In general this illustration has important policy overtones, for adopting different techniques or inventing new technologies to change the coefficients in the mass flow-consumer services relationship may be easier than accommodating the economy to extensive recycling. At least, such an alternative should not be disregarded. Just as changing production technology is incorporated into the Walras-Cassel model by adding alternative production functions for each good, so should new or existing alternatives in the technology of consumption be incorporated by adding equations to the model representing the relationships between goods and goods characteristics actually demanded.

The extended Walras-Cassel model includes a new set of "resources" which provide undesired inputs to the production of the final demands. Ayres and Kneese identify these resources as the pollutants, and associate a negative price with each. The key relations defining the role of these resources are a production function:

$$(2) \quad r_k^u = G_k^u(Y_1, \ldots, Y_n) \qquad k = 1, \ldots, S$$

where r_k^u is the kth undesirable resource and Y_i is the ith final good, and a forced consumption function:

$$(3) \quad r_k^u = \sum_{j=1}^{N} b_{kj}^u Y_j \qquad k = 1, \ldots, S$$

where b_{kj}^u is the amount of resource r_k^u entering, directly or indirectly (and undesirably), into the production of Y_i. The Walras-Cassel model is then solved for a set of prices which is consistent with the final demands and the resource availabilities. Ayres and Kneese assert, in their formulation, that the extended Walras-Cassel model is useful for calculating the prices associated with the undesired inputs, and that these prices will be negative. Unfortunately, equation (3) will *not* produce this result. The problem is that nowhere in the Ayres-Kneese model does pollution generate costs, either as disutility to consumers or additional resource requirements to producers. Whether a joint product will have a positive, zero, or negative price depends upon the level of output (from (2)), its usefulness as an input (from (3)), its usefulness as a final demand, and the resource costs of offsetting its effects or disposing of it. The last two features are excluded from the Ayres-Kneese model; to include them requires several alterations.

First, some "final demands" are for the absence of pollution: air free of eye irritants and permitting visibility, water pure enough for fish to survive and people to swim, etc. These final demands are often discontinuous functions of the level of pollution, acting as a demand for a constraint on pollutants. Corresponding to the preceding examples, marginal reductions in pollutants in a badly polluted river do not make it marginally less unsuitable for swimming or fishing, nor is the limitation on the level of pollution required for eye irritation to be avoided proportional to the amount of seeing (i.e., population).[4]

If pollution abatement can be included in the list of final demands, the Walras-Cassel model can reflect an important characteristic of the actual pollution problem: competing final demands are sometimes mutually inconsistent. For example, a river cannot be used as both an open sewer and a public recreational resource; no marginal adjustments of benefits and costs of recreation and sewage transportation are possible once pollution has exceeded a certain level far below the maximum transport capacity of the river. The mathematical programming approach to general equilibrium solution (maximizing gross national output) can be used in these cases to estimate the cost of obeying an antipollution constraint, or conversely of polluting sufficiently heavily to preclude recreational uses.

---

[4] By similar argument, these are not demands for incremental units of a common property resource, but that the whole resource, regardless of other uses, have certain minimal physical properties.

The second necessary alteration of the Ayes-Kneese general equilibrium model arises from the fact that equation (3) does not adequately represent the way in which pollutants affect production processes, for its form is identical to equations expressing production resource requirements. In fact, Ayres and Kneese have extended the Walras-Cassel model to include joint products, in general, *not* undesirable inputs specifically. Equation (3) is identical in form to the resource requirement equations, and also to a linear version of the production functions for residuals, (2).

That undesirable resources can have positive prices in the Ayres-Kneese model is easy to demonstrate. The equation (3) can be interpreted to mean: "If Y_w is the final demand for potable water, the total amount of pollutant k in water for drinking, r_k^u, must be $b_{kw}^u Y_w$." Suppose that for a set of final demands $\{Y^0\}$ the equation (3) will hold. Suppose we wish to determine the effect of moving to a new set of final demands, $\{Y^1\}$, in which $Y_i^1 = Y_i^0$ except for $i=w$, in which case $Y_w^1 = Y_w^0 + dy_w$. Since the production of Y_w does not produce undesirable resource r_k^u, this new final demand cannot be met—pollutant r_k^u is a "scarce" resource, unable to meet the "demand" (3), and its price will therefore be positive.

The difficulty, of course, is that the coefficient b_{kj}^u is not constant, but depends upon the level of pollution and, because of (2), on the level and mix of final demand.

Pollutants should enter a general equilibrium model as factors generating additional resource and service requirements in production. Thus the equations representing resource and service use in production might appear as:

$$t_j = \sum_i \left(a_{ij} + \sum_k b_{ijk} r_k^u\right) Y_i \tag{4}$$

$$j = 1, \ldots, M; M = L + P$$

where t_j is the use of factor j, b_{ijk} is the additional amount of factor j needed to produce a unit of final good i if pollutant k increases by one unit, and r_k^u is the level of pollution of pollutant k.[5] Equation (4) easily permits inclusion of pollution abatement as final demand. The Y_i refers to the degree of final purity desired, and the term in brackets the resource requirement for achieving abatement (as a function of initial pollution).

If the extended Walras-Cassel model is amended by including equation (4) as representative of the resource requirements equations, the market equilibrium equations must also be altered:

$$P_i = \sum_j V_j \left(a_{ij} + \sum_k b_{ijk} r_k^u\right) \tag{5}$$

$$i = 1, \ldots, N$$

or, if the creator of pollution is to pay,

$$P_i = \sum_j V_j a_{ij} + \sum_k V_k^u d_{ik}^u \tag{5'}$$

$$i = 1, \ldots, N$$

where P_i, V_j, and V_k^u are the prices of final good i, desirable resource j, and undesirable resource k, and d_{ik}^u is the output of pollutant k resulting from the production of a unit of final good i. To make pollution damage equal to revenues, a final "market equilibrium" expression is needed:

$$V_k^u = \sum V_\imath \left(\sum_i b_{ijk} Y_i\right) \tag{6}$$

$$k = 1, \ldots, S$$

---

[5] Two points about equation (4) should be noted. First, b_{ijk} is the product of two terms: the fraction of pollutant k affecting the production of final good i and the additional amount of resource j required in the production of i if the production of good i is subjected to one more unit of pollutant k. Second, the t's in (4) include the resources r and services s that appear in equation (1)..

Equations (4), (5), and (6) are non-linear, so that investigating mathematical conditions for equilibrium is more difficult than in the Ayres-Kneese model. Unfortunately, equations (4) and (6) cannot be made linear without extremely heavy losses in the usefulness in the model. To make (4) and (6) linear requires one of two assumptions: (1) the costs of a given level of pollution are independent of the size and mix of final demand,[6] or (2) the costs of pollution abatement are unrelated to the level of pollution.

III. Conclusions

The preceding sections contain several proposals for generalizing the Ayres-Kneese model to make it more realistic and more applicable to pollution policy. Ayres and Kneese developed their model with policy applications in view; they hope to see a general equilibrium approach used for projecting residual deposition and for evaluating alternative control techniques. Before these aspirations become realities, a number of tough theoretical and empirical problems must be solved. This paper has suggested four important extensions not found explicitly in the Ayres-Kneese model: 1) separating "residuals" from "pollutants" and inclusion of the complex relations between the two categories (much of which is lost through the mass balance approach that neglects differences in the seriousness of different types of pollutants and that ignores interactions among residuals and pollutants); 2) including pollution abatement as a final demand, sometimes in the form of a collective good and as a constraint on the production system; 3) freeing the fixed relationship between goods and consumer services by recognizing that in consumption, like production, opportunities exist for switching to different methods of producing goods characteristics; and 4) correcting the equation representing the effect of pollution on production to avoid the necessity for pollution as an input that is implicit in the Ayres-Kneese model. While accounting for all of these realities necessarily complicates the general equilibrium model, such complexities are necessary if the model is to be relevant to pollution abatement planning.

REFERENCES

R. V. Ayres and A. V. Kneese, "Production, Consumption and Externalities," *Amer. Econ. Rev.*, June 1969, *59*, 282–97.

A. O. Converse, "On the Extension of Input-Output Analysis to Account for Environmental Externalities," *Amer. Econ. Rev.*, Mar. 1971, *61*, 197–98.

K. Lancaster, "A New Approach to Consumer Theory," *J. Polit. Econ.*, Apr. 1966, *74*, 132–57.

———, "Change and Innovation in the Technology of Consumption," *Amer. Econ. Rev. Proc.*, May 1966, *56*, 14–23.

W. Leontief, "Environmental Repercussions and the Economic Structure: An Input-Output Approach," *Rev. Econ. Statist.*, Aug. 1970, *52*, 262–71.

G. Stigler, "Production and Distribution in the Short Run," *J. Polit. Econ.*, June 1939, *48*, 305–27.

---

[6] Recently Wassily Leontief has incorporated environmental pollutants into his input-output model by making this assumption. Leontief's model includes the production functions for direct pollution abatement, but does not permit pollutants to alter resource requirements in the production of other goods.

Externalities as Commodities

By Robert A. Meyer, Jr.*

The development of general equilibrium theory has proceeded with a wide variety of alternative sets of assumptions, but with at least one particular assumption common to all—the absence of externalities on both the production and consumption sides of economic activity. A. C. Pigou in *The Economics of Welfare* pointed out the now well-known inefficiency which may emerge as the result of a perfectly competitive market environment in the presence of externalities. The birth of a new conscience with respect to environmental problems, particularly those related to economic activity, has led to a new interest in a well-known proposal for handling externalities—taxes and subsidies. In a recent article in this *Review*, F. Trenery Dolbear, Jr. provided a geometric argument which led him to conclude that the "... analysis suggests that a per unit tax ... which has the twin objectives of keeping [an individual] on his starting point indifference curve and attaining a Pareto optimum is not, in general, possible" (p. 98).

This paper constructs a general equilibrium model with externalities in non-linear programming terms, and it provides sufficient conditions for existence of a solution. Traditional general equilibrium models may be interpreted as a special case. The center of interest is the economic interpretation of the dual solution which yields the well-known interpretation of the Lagrange multipliers as prices for commodities, moreover, it provides a set of variables associated with "externality technologies" which may be interpreted as taxes and subsidies. The result obtained may be taken as further evidence of the fact that it is the non-convexities, not the externalities per se, that raise difficulties.

The particular class of admissible tax-subsidy schemes which has received frequent attention from economists is the one in which the tax or subsidy is a constant amount per unit. A tax-subsidy scheme of this type will be referred to as a Pigouvian style tax-subsidy scheme. For questions dealing with externalities, this means that the set of admissible pricing schemes is of the same type as the set of admissible commodity price schemes that is employed in general equilibrium analyses where one rules out the admissibility of prices which vary over inframarginal units. This similarity may be exploited more fully by treating externalities as additional commodities and then establishing the existence of "prices" associated with them in an extended commodity space.

The negative conclusion of Dolbear raises at least two questions: 1) when can a Pareto optimal point be achieved as an equilibrium (in the presence of externalities) by the use of the Pigouvian style tax-subsidy scheme, and 2) when will the constant amount per unit tax-subsidy scheme simultaneously achieve a Pareto optimum and yield exact compensation (a point complementary to Dolbear's conclusion).

In the context of externality discussions, as in the original discussions of compensation tests by J. R. Hicks, N. Kaldor, T. Scitovsky, and I. M. D. Little, an initial state must be specified. It has been emphasized by Ronald Coase and Harold Demsetz that for questions of externalities both the initial condition and the set of admissible alternatives is delineated by the legal environment which one assumes. Dolbear adopted two polar cases as a basis for discussion that are analogous to pure competition and pure monopoly. In one case, no recourse is available to an injured party (for Dolbear this example is smoke pollution; one individual consumes bread and smoke, the other bread and heat which generates smoke). In the other case, no injury is permitted except with the consent of the injured party. The latter defines a reference point such that no alternative solution is admissible unless it is at least as preferred as the initial state. Exact compensation simply restricts one to alternative states which are indifferent to the initial state.

---

* Assistant professor of economics, Purdue University. I would like to express my appreciation for the helpful comments of Charles R. Plott, Andrew B. Whinston, and James C. Moore as well as a referee and Martin Ringo, whose suggestions measurably improved the exposition.

There are several points on which Dolbear is not specific:

Frequent reference is made to market prices for his two-commodity world but there is no mention of how these are determined, or even if they can be shown to exist.

Even if prices for bread and heat exist, decentralized production by profit maximizers may not lead to a unique production point since a linear production transformation frontier is assumed.

The assumptions about consumers is that they have indifference curves which . . . "meet the usual requirements in the theory of consumer choice" (p. 91).

To provide at least one possible alternative development which will allow one to obtain further results, a non-linear programming framework will be employed which has the immediate consequence of providing one with a well-developed body of theory to call upon. In addition, a different method of handling externalities is employed which will allow explicit representation of technologies which mitigate the amount of an externality. While this may appear somewhat awkward for some problems, its interpretation as smoke abatement technology for the case Dolbear considered seems quite natural. Although the altered production possibilities set obtained by extension to include each externality as a separate commodity is developed here, as in Dolbear, for externalities in consumption, the representation of externalities in production in an analogous form is also possible.

The negative conclusion of Dolbear is presumably predicated on some type of non-convexity which is not stated explicitly. It is shown here that the absence of non-convexities is the prime assumption which yields a conclusion contrary to Dolbear's.

As a setting for a general equilibrium problem with externalities treated as additional commodities via "technologies" let us assume an economy with the following characteristics:

There are a finite number of individuals, $i=1, \ldots, m$, and each person's preferences are representable[1] by a quasic-concave utility function, U^i.

There are a finite number of primary commodities denoted $X_j, j=1, \ldots, n$, where the amount of the jth commodity consumed by the ith individual is denoted x_j^i.

Externalities are assumed to arise as a result of interdependence of preferences over consumption of primary commodities, X_j. Each consumption gives rise to a vector of "externality commodities" of the form $x_{jk}^i = g^i(x_j^k)$, where i and k refer to individuals and j to primary commodities; $i=1, \ldots, m$; $j=1, \ldots, n$; and $k=1, \ldots, m$, except $k \neq i$. For $n=2$, $m=2$ we have the case analyzed by Dolbear where g^i is the transformation which links quantities of heat produced with quantities of smoke. Each person's utility function has the arguments

$$U^i(\overset{i}{x}_1, \ldots, \overset{i}{x}_n, \overset{i}{x}_{1,1}, \ldots, \overset{i}{x}_{1,i-1}, \overset{i}{x}_{1,i+1}, \ldots, \overset{i}{x}_{n,m})$$

The economy is assumed to be confronted by a combined resource-technological knowledge constraint which defines a convex production possibilities set. It will be convenient to assume that the feasible production set can be represented by the set satisfying $X \geq 0$ and $F(X) \leq 0$ where F is strictly quasi-concave,[2] $X \equiv (X_1, \ldots, X_n)$.

Since we are concerned with Pareto optimal solutions, the welfare economics nature of the problem leads to the formulation of the general equilibrium problem as:[3]

---

[1] These conditions are clearly much more stringent than necessary, however, they do allow a simpler argument in non-linear programming terms rather than using set theoretic terms.

[2] The comment of fn. 1 applies here as well.

[3] An equivalent formulation would be a "quasi-concave vector maximum" problem of the form: max $G(Z)$, an m-vector, subject to $H(Z) \geqq 0$ and $Z \geqq 0$, where we define

$$\begin{aligned} Z &\equiv (z_1, \ldots, z_{n+nm^2})' \\ &\equiv (X_1, \ldots, X_n, \overset{1}{x}_1, \ldots, \overset{m}{x}_n, \overset{1}{x}_{12}, \ldots, \overset{m}{x}_{n,m-1})' \\ G(Z) &\equiv (G_1(Z), \ldots, G_m(Z))' \equiv (U^1, \ldots, U^m)' \end{aligned}$$

$$\text{(1)} \qquad \max U^m$$

subject to

$$\text{(2)} \qquad \sum_{i=1}^{m} \overset{i}{x}_j \leq X_j \qquad j = 1, \ldots, n$$

$$\text{(3)} \qquad F(X) \leq 0$$

$$\text{(4)} \qquad X_j \geq 0, \quad \overset{i}{x}_j \geq 0, \quad \overset{i}{x}_{jk} \geq 0$$

$$\text{(5)} \qquad U^i \geq \overline{U}^i \qquad i \neq m$$

$$\text{(6)} \qquad \overset{i}{x}_{jk} = \overset{i}{g}(\overset{k}{x}_j) \qquad \begin{array}{l} i = 1, \ldots, m \\ j = 1, \ldots, n \\ k = 1, \ldots, m, \\ \quad \text{except } k \neq i \end{array}$$

If the externality technologies, g^i, are all identically zero, then the above problem is a non-linear programming formulation of the problem dealt with in modern general equilibrium discussions. An example of this method may be found in Otto Davis and Andrew B. Whinston. If the constraints, (6) are present then a set of sufficient conditions for solution of this problem will yield a set of dual variables which include not only the traditional competitive prices for the n primary commodities but in addition a set of "prices" which may be interpreted as the associated constant amount per unit taxes or subsidies on the externalities.[4]

Since the utility functions are assumed quasi-concave and each of the constraints is quasi-concave one may use the theorems developed by Kenneth J. Arrow and Alain C. Enthoven to provide sufficient conditions under which the Kuhn-Tucker conditions[5] associated with (1)–(6) will be sufficient to characterize (and imply existence of) a solution. If we assume: 1) that the Kuhn-Tucker conditions are satisfied for some values of X_j, $\overset{i}{x}_j$, and $\overset{i}{x}_{jk}$ and 2) that there is at least one argument of U^m, say x_i^m, such that $\partial U^m / \partial x_i^m > 0$ when evaluated at the solution values, then by Arrow-Enthoven Theorem 1, these are sufficient to establish existence. Assumption 2 is another way of saying that individual m is not satiated at the optimum point. This condition is assured in traditional arguments by the assumption of no bliss points for consumers.

Having established existence of a solution, the prime interest lies in the vector of Lagrange multipliers associated with the constraints. The first n components of the vector of optimal values for the multipliers, λ^0, may be interpreted as the competitive equilibrium prices associated with the primary commodities. The $n+1$ component is

---

where a prime denotes the transpose.

$$H(Z) \equiv \begin{bmatrix} X_j - \overset{i}{x}_j \\ -F(X) \\ \overset{i}{x}_{jk} - \overset{i}{g}(\overset{k}{x}_j) \\ \overset{i}{g}(\overset{k}{x}_j) - \overset{i}{x}_{jk} \end{bmatrix} \geqq 0 \qquad \begin{array}{l} j = 1, \ldots, n \\ i = 1, \ldots, m \\ k = 1, \ldots, m \\ i \neq k \end{array}$$

Then we have the following theorem (Theorems 4 and 5 from H. W. Kuhn and A. W. Tucker):

> If Z^o is a proper solution to the vector maximum problem (i.e., there are no other Z's satisfying the constraint set such that $G_i(Z) \geqq G_i(Z^o)$ for all i and $G_i(Z) > G_i(Z^o)$ for some i), then there exists some $n+1+2[nm(m-1)]$-vector, $v^o > 0$, such that for Z^o and some $\lambda^o \geqq 0$

$$\text{(1a)} \quad \left(\frac{\partial \phi}{\partial z_i}\right)_{Z=Z^o} \leqq 0, \quad \left(\frac{\partial \phi}{\partial z_i}\right)'_{Z=Z^o} \cdot Z^\circ = 0, \quad Z^\circ \geqq 0$$

$$\text{(2a)} \quad \left(\frac{\partial \phi}{\partial \lambda_i}\right)_{\lambda=\lambda^o} \geqq 0, \quad \left(\frac{\partial \phi}{\partial \lambda_i}\right)'_{\lambda=\lambda^o} \cdot \lambda^\circ = 0, \quad \lambda \geqq 0$$

where

$$\phi \equiv \phi(Z, \lambda) \equiv v^{o\prime} G(Z) + \lambda' H(Z) \quad \text{and} \quad \left(\frac{\partial \phi}{\partial z_i}\right)_{Z=Z^o}$$

is an $(n+nm^2)$-vector of partial derivatives evaluated at $Z = Z^o$. If we add to the necessary conditions that

$$\text{(3a)} \quad \begin{array}{l} \phi(Z, \lambda^\circ) \leqq \phi(Z^\circ, \lambda^\circ) + \left(\dfrac{\partial \phi}{\partial z_i}\right)'_{Z=Z^o} \cdot (Z - Z^\circ) \\ Z \geqq 0 \end{array}$$

it is then sufficient for Z^o to be a proper solution. The set of all Pareto optimal solutions is the set of solutions to a problem phrased as a vector maximum problem since the definitions of a vector maximum and a Pareto optimum coincide. Formulation as a vector maximum problem is an alternative way of developing sufficient conditions for the existence of a solution.

[4] If one were going to use the interpretation suggested here to apply to externalities in production, or even in the case of only externalities in consumption, the possibility of inferior inputs discussed by Charles R. Plott raises difficulties as to whether an output should be taxed (e.g., heat in Dolbear's case) or the use of inputs should be taxed.

[5] In the notation used by Kuhn and Tucker, (1a)–(3a) of fn. 3 are the Kuhn-Tucker conditions.

the marginal value[6] of the combined resource-technological knowledge constraint on production of primary commodities. These interpretations are familiar.

The presence of externalities does not alter the usual interpretation above but appends to it an additional set of $mn(m-1)$ Lagrange multipliers associated with constraints (6). Constraints (2) and (3) were inequality constraints which implied nonnegative multipliers associated with them. For constraints (6) this is not true. The equality constraints of the externality technologies carry with them an implicit nonrejectability property (which has been discussed by the author elsewhere in a public goods context) and also the possibility that the externalities may be "goods" ($\lambda > 0$) as well as "bads" ($\lambda < 0$). Each multiplier may be interpreted in a manner analogous to that used for the primary commodities as a price associated with each particular externality. These may also be interpreted as the constant amount per unit taxes or subsidies of a Pigouvian type. Interpreted in this way the sufficient conditions given above answer the first question by providing a set of conditions under which a Pareto optimal point, even in the presence of externalities, can be achieved by an appropriately specified set of prices.

While it is true that one may use a tax-subsidy interpretation, there is nothing here to imply who should pay whom. A tax paid by i to j is a tax from i's standpoint but a subsidy from j's. This problem is not new to welfare economics discussions, having also played a rôle in the compensation test discussions of Hicks and Kaldor.

The prices associated with the externalities are not of course prices established in the same sense that prices are formed in a competitive market since they are prices in a "market" with a single buyer and a single seller. This does not detract from the fact that an individual consumer's behavior, modeled as utility maximization subject to a budget constraint utilizing the prices given by the optimum value of the Lagrange multipliers, will lead to an equilibrium which is Pareto optimal in the same manner as shown by Davis and Whinston. The apparent absence of markets in external effects and the inference that this implies a lack of Pareto optimality has been attacked by Harold Demsetz who argues that such an inference presupposes that the formation and operation of such markets utilizes no resources. In effect, Demsetz treats markets and institutions themselves as commodities in the same way externalities have been treated here.

To answer the second question concerning existence of an equilibrium in the exact compensation case, one need only replace constraints (5) by equality constraints of the form $U^i = \overline{U}^i$, $i = 1, \ldots, r$, $i \neq m$, for those individuals for whom exact compensation is to be achieved. For the remaining $m-r-1$ individuals, the constraint would remain as it is in (5). The same conditions given above provide sufficient conditions for the exact compensation case.

In the model used in this paper, the externality transformation functions, g^i, were taken as given; however, in many externality problems such as environmental pollution control the g^i may themselves be treated as variables for analytical purposes. To treat the g^i as variables means one may use resources to alter the present state of technology as well as to produce commodities for current consumption. Such a treatment is analogous to making technological change endogenous in a growth model. As an example one might think of the Dolbear model, in that case choice among alternative g^i would mean choice among alternative smoke emission control techniques.

REFERENCES

K. J. Arrow and A. C. Enthoven, "Quasi-Concave Programming," *Econometrica*, Oct. 1961, *4*, 779–800.

M. L. Balinski and W. J. Baumol, "The Dual in Nonlinear Programming and its Economic

---

[6] The validity of interpreting the multipliers in a non-linear program as marginal valuations is dependent on appropriate differentiability conditions prevailing at the solution point. For a detailed discussion of this point the reader is referred to M. L. Balinski and William J. Baumol.

Interpretation," *Rev. Econ. Stud.*, July 1968, *35*, 237–56.

J. M. Buchanan and W. C. Stubblebine, "Externality," *Economica*, Nov. 1962, *29*, 371–84.

R. Coase, "The Problem of Social Cost," *J. Law and Econ.*, Oct. 1960, *3*, 1–44.

O. Davis and A. B. Whinston, "Some Notes on Equating Private and Social Cost," *Southern Econ. J.*, Oct. 1965, *32*, 113–25.

——— and ———, "Welfare Economics and the Theory of the Second Best," *Rev. Econ. Stud.*, Jan. 1965, *32*, 1–14.

H. Demsetz, "The Exchange and Enforcement of Property Rights," *J. Law Econ.*, Oct. 1964, *7*, 11–26.

F. T. Dolbear, Jr., "On The Theory of Optimum Externality," *Amer. Econ. Rev.*, Mar. 1967, *57*, 90–103.

J. R. Hicks, "Foundations of Welfare Economics," *Econ. J.*, Dec. 1939, *49*, 696–712.

N. Kaldor, "Welfare Propositions of Economics and Interpersonal Comparisons of Utility," *Econ. J.*, Sept. 1939, *49*, 549–52.

H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in J. Neyman, ed., *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley 1951.

I. M. D. Little, *A Critique of Welfare Economics*, 2d ed., Oxford 1957.

J. E. Meade, "External Economies and Diseconomies in a Competitive Situation," *Econ. J.*, Mar. 1952, *62*, 54–67.

R. A. Meyer, Jr., "Private Costs of Using Public Goods," *Southern Econ. J.*, Apr. 1971, *38*, 479–90.

A. C. Pigou, *The Economics of Welfare*, 4th edition, London 1932.

C. R. Plott, "Externalities and Corrective Taxes," *Economica*, Feb. 1966, *33*, 88–91.

T. Scitovsky, "A Note on Welfare Propositions in Economics," *Rev. Econ. Stud.*, Nov. 1941, *9*, 77–88.

R. Turvey, "On Divergences Between Social Cost and Private Cost," *Economica*, Aug. 1963, *30*, 309–13.

Stochastic Reserve Losses and Expansion of Bank Credit: Note

By J. PHILLIP COOPER*

In 1961, in this *Review*, Daniel Orr and W. G. Mellon developed a model which demonstrates the optimal credit expansion for a profit-maximizing bank facing random cash flows and a legal reserve requirement. Comparisons are drawn between the bank's response to an injection of excess reserves under conditions of uncertainty and the action indicated by traditional deposit-creation analysis. The primary purpose of this note is to point out a major error in the presentation of this model, the misleading parametric tabulation of the optimal solution. This problem is not specific to the Orr and Mellon model, but is common in work requiring numerical examples. For this reason, two other logical errors—misspecification of the bank's expected profit function and consideration of only local optima—are relegated to footnotes.[1] No attempt is made here to update the model to hold under current reserve requirement regulations.

The reader is referred to the original article for a full discussion of the model and assumptions. For convenience, the variables are defined below:

$R \equiv$ the volume of excess reserves at the beginning of the evaluation period
$D \equiv$ the volume of new deposit liabilities created during the period via the granting of loans
$L \equiv$ the loss of reserves during the period
$\rho \equiv$ the legal reserve ratio $(0 < \rho < 1)$
$i \equiv$ rate of interest earned on loans
$M \equiv$ lump sum penalty for violation of reserve requirements
$r \equiv$ penalty for each dollar of reserves which the bank is short
$P \equiv$ expected profit over the evaluation period
$f(L) \equiv$ probability density function of L (assumed normal with mean proportional to the level of new deposits created during the period but variance independent of deposit liabilities)
$k \equiv$ ratio of mean reserve loss to level of new deposits created

The model can be summarized as follows:

Reserve Loss

$$(1) \quad L = kD + \epsilon \quad \text{with} \quad \epsilon \sim N(0, 1)$$

Probability Density Function

$$(2) \quad f(L) = N(kD, 1) = \frac{1}{\sqrt{2\pi}} e^{-((L-kD)^2/2)}$$

Reserve Condition

Amount of deficiency

$$(3) \quad \begin{aligned} &= \text{Reserves required} - \text{Reserves held} \\ &= \quad \rho(D - L) \quad - \quad (R - L) \\ &= (1-\rho)L - (R - \rho D) \\ &= (1-\rho)\left(L - \frac{R - \rho D}{1-\rho}\right) \\ &= (1-\rho)(L - v) \end{aligned}$$

where

$$v \equiv \frac{R - \rho D}{1 - \rho}$$

The term v is the critical value of actual reserve losses above which a deficiency will be realized. Note that each dollar of reserve loss in excess of this critical value leads to a deficiency of only $(1-\rho)$ dollars (because of the reserves "freed up" when deposits are withdrawn).

---

* Instructor in business economics and finance at the University of Chicago, Graduate School of Business.

[1] The original article also contained a differentiation error, subsequently corrected by the authors in Errata. Because of the first and third errors mentioned in the text, the authors were unable to see any significant change in the corrected optimal solutions.

Expected Profit Function

$$P(D) = iD - M\int_v^\infty f(L)dL - r\int_v^\infty (1-\rho)(L-v)f(L)dL \tag{4}$$

The first term represents the interest earned on loans granted during the period; the second term is the lump sum penalty times the probability of reserve deficiency; and the third term is variable penalty cost times the expected deficiency.[2]

A necessary condition for the optimality of a lending policy is that[3]

$$\frac{\partial P}{\partial D} = i - \left(\frac{\rho}{1-\rho}+k\right)\{Mf(v) + r(1-\rho)[1-F(v)]\} = 0 \tag{5}$$

where

$$F(x) \equiv \int_{-\infty}^{x} f(L)dL$$

Equation (5) can be solved by numerica methods for the value of D, say $\hat{D}$. Thi solution is a local maximum only if

$$\frac{\partial^2 P}{\partial D^2} = -\left(\frac{\rho}{1-\rho}+k\right)^2 f(\hat{v})[M(\hat{v}-k\hat{D}) + r(1-\rho)] < 0 \tag{6}$$

where $\hat{v}$ is the value of v for $D=\hat{D}$.

The slope of the expected profit functio given in (5) approaches a limit asymptotically for indefinitely large values of D.

$$\begin{aligned} \lim_{D\to\infty}\left(\frac{\partial P}{\partial D}\right) &= i - \left(\frac{\rho}{1-\rho}+k\right)[r(1-\rho)] \\ &= i - [\rho + k(1-\rho)]r \end{aligned} \tag{7}$$

Asymptotically, one dollar in credit expansions earns i dollars. It will also increase th already certain deficiency by requiring dollars more in reserves and by producin additional reserve losses of k dollars or a extra reserve shortage of $(1-\rho)k$ dollars this total asymptotic deficiency o $[(1-\rho)k+\rho]$ dollars incurs a penalty at rat r. If

$$\lim_{D\to\infty}\left(\frac{\partial P}{\partial D}\right) > 0,$$

then it should be recognized that, even if th solution of (5) is a local maximum, it will no be the global maximum.[4] If, on the othe

---

[2] At this point, the first additional shortcoming of the Orr and Mellon paper has been encountered. Equation (4) was derived logically from the assumptions of the model and is in complete agreement with the Orr and Mellon text. Their own mathematical specification of the expected profit function (equations (2) and (4), p. 617) differs, however, in that each dollar of reserve loss, rather than reserve deficiency, is penalized.

[3] Rewrite (4) as

$$P = iD - [M - rv(1-\rho)]\int_v^\infty f(L)dL - r(1-\rho)\int_v^\infty Lf(L)dL;$$

$$\frac{\partial P}{\partial D} = i - [M - rv(1-\rho)]f(v)\left[\frac{\rho}{1-\rho}+k\right] - [1-F(v)](r\rho) - r(1-\rho)vf(v)\left(\frac{\rho}{1-\rho}\right) - r(1-\rho)\int_v^\infty L\frac{\partial f(L)}{\partial D}dL$$

Integrate by parts, as follows:

$$\begin{aligned} \int_v^\infty L\frac{\partial f(L)}{\partial D}dL &= \int_v^\infty Lf(L)(L-kD)kdL \\ &= k[-Lf(L)]_v^\infty - k\int_v^\infty -f(L)dL \\ &= k\{vf(v) + [1-F(v)]\} \end{aligned}$$

and substitute in the expression for $\partial P/\partial D$. Thus,

$$\begin{aligned} \frac{\partial P}{\partial D} &= i - f(v)\left(\frac{\rho}{1-\rho}+k\right) \cdot \{[M - rv(1-\rho)] + r(1-\rho)v\} - [1-F(v)][r(1-\rho)]\left(\frac{\rho}{1-\rho}+k\right) \\ &= i - \left(\frac{\rho}{1-\rho}+k\right)\{Mf(v) + r(1-\rho)[1-F(v)]\} \end{aligned}$$

[4] This is the second additional shortcoming of th Orr and Mellon paper; in fact, most of their Section I requires restatement. While the infinite solution ma not be realistic, the model must be modified to preclud it, say by maximizing expected profits subject to a upper-bound constraint on the probability of deficiency. If we define the probability of a deficiency

Table 1—Optimal Values of D under Uncertainty Compared with Traditional Values

Case	M	r	i	k	R	D_{err}	D_{cor}	α	D_{mod}	α	D^*
1	20	.01	.0025	.7	10	8.9	8.9	.0000	8.9	.0000	13.2
2	20	.01	.0025	.7	1000	1311.	1311.	.0000	1312.[a]	.0000	1316.
3	0	.01	.0025	.7	10	10.5	10.7	.0097	12.7	.3289	13.2
4	20	.01	.0025	.3	10	15.7	15.7	.0001	15.7	.0001	22.8
5	20	.5	.0025	.7	10	8.9	8.9	.0000	8.9	.0000	13.2
7	20	.01	.0025	0	10	35.4	∞	1.0000	∞	1.0000	50.

[a] In case 2, $D_{\text{mod}}-D_{\text{cor}}=0.2$ although this produced a change in the fourth significant digit.

hand, the limit is negative, no infinite solution exists. Note also that

$$\lim_{D\to-\infty}\left(\frac{\partial P}{\partial D}\right)=i>0 \tag{8}$$

Infinite contraction is never profitable.

Table 1 is an expansion of its counterpart in the Orr and Mellon article, p. 619; the case numbers and parameter values are retained; ρ is set at .2.[5] Proceeding from left to right in the table, the following additional definitions are necessary:

$D_{\text{err}}\equiv$original Orr and Mellon solution

$D_{\text{cor}}\equiv$solution to Orr and Mellon objective function, differentiation corrected[6]

$\alpha\equiv$probability of deficiency, i.e., $\int_v^\infty f(L)dL$

$D_{\text{mod}}\equiv$solution to objective function of equation (4), i.e., $\hat{D}$

$D^*\equiv$deterministic solution, the "traditional" value[7]

The value of D^* is calculated by suppressing the stochastic disturbance ϵ in (1), that is, by assuming

$$L=kD \tag{1'}$$

and by finding the value of L that would just exhaust the initial excess reserves, R; that is,

$$L=v \tag{9}$$

Equations (9) and (1′) together give

$$D^*=\frac{R/(1-\rho)}{\left(\dfrac{\rho}{1-\rho}+k\right)} \tag{10}$$

Introducing the concept α sheds light on

$$\alpha\equiv\int_v^\infty f(L)dL=1-F(v),$$

we can see that

$$\frac{\partial\alpha}{\partial D}=f(v)\left(\frac{\rho}{1-\rho}+k\right)>0$$

This implies that we can find P as a function of α, $P(\alpha)$, because of the strictly monotonic relationship between α and D. The probability of a deficiency and the expected profit at the local optimal solution are $\alpha(\hat{D})$ and $P[\alpha(\hat{D})]$, respectively. Consider the segment of $P(\alpha)$ which is rising without limit as α approaches unity and define $\tilde{\alpha}$ as that value of α which satisfies $P(\tilde{\alpha})=P[\alpha(\hat{D})]$. Then, any upper bound on α falling between $\alpha(\hat{D})$ and $\tilde{\alpha}$ is sufficient to reestablish $\hat{D}$ as the valid optimal solution. (It may be shown that $P(\alpha)$ has at most two stationary points.) $\tilde{\alpha}$ turned out to be very large in the many cases tabulated in preparing this paper; since, at this point, many of the assumptions of the model become especially strained (e.g., r, i, and variance of L all independent of D) this variation will not be pressed further except to note that we could also constrain the expected reserve deficiency. In this simple model, the two constraints have a one-to-one correspondence. The first one can be represented as:

$$\alpha=1-\psi(v-kD)<\beta$$

where ψ is the distribution function of a standard normal variate and β is the maximum acceptable risk of deficiency. The alternative can be represented as: expected deficiency$=(1-\rho)(kD-v)<\gamma$ where γ is the maximum acceptable deficiency in dollars. Both imply constraints on $(v-kD)$ which is monotonic in D.

[5] The relationship of r to i is puzzling; Orr and Mellon offer no explanation.

[6] The limit (7) in the original article was $i-r$; with the differentiation corrected, the limit was $i-kr$. The authors did not recognize the impact of their correction on the solution for case 7. Case 6 is also irrelevant because of the differentiation error.

[7] In case 7, D^* was incorrectly calculated at 40 in the original article. This number was also used in their text, p. 621. The Errata noted this error.

whether or not the authors' tabulation of optimal values of D has any validity. First, the values of α, not calculated by Orr and Mellon, are all ridiculously low for the finite solutions. Second, with the exception of the infinite solution in case 7, $D_{\text{err}} \approx D_{\text{cor}}$. Third, even with the objective function presented here, D_{mod} differs significantly from D_{cor} in only one case, the third. Not coincidentally, this is the only finite solution to exhibit a nontrivial probability of ever being deficient. The specific characteristic of this case is the zero lump sum penalty cost.

The problem here is that the value chosen for one of the parameters is inappropriate for the assumptions of the model. When the variance of reserve losses is set at unity, the system is no longer scale free. Taking M to have a value of 20 turns out to be almost equivalent to imposing an indefinitely large lump sum penalty on the realization of a deficiency. One ramification of this large fine was the surprising result that neither mathematical error nor misspecification produced any change in the finite optimal solutions. Table 2 is presented below to show the sensitivity of $\hat{D}$, α, and P to M; i is set at .0025, k at .7, and R at 10.

TABLE 2—SENSITIVITY OF $\hat{D}$, α, AND P TO CHANGES IN M

r	M	ρ	$\hat{D}$	α	P
.01	0.0	.1	13.20	.342	.0309
		.2	12.69	.329	.0300
	0.1	.1	10.89	.011	.0261
		.2	10.69	.010	.0257
	1.0	.1	9.85	.001	.0237
		.2	9.82	.001	.0238
.005	0.0	.1	14.29[a]	.685	.0326
		.2	13.59[a]	.658	.0314
	0.1	.1	10.90	.012	.0261
		.2	10.70	.010	.0258

[a] Expansion here greater than under certainty.

TABLE 3—SENSITIVITY OF $\hat{D}$ TO PARAMETERS WHEN $R=1.0$

r	M	ρ	$\hat{D}$	α	P
.01	0.0	.1	.869	.342	.0001
		.2	.850	.329	.0004
	0.1	.1	−1.433	.011	−.0048
		.2	−1.152	.010	−.0039
	1.0	.1	−2.48	.001	−.0071
		.2	−2.02	.001	−.0058
.005	0.0	.1	1.964	.685	.0018
		.2	1.744	.658	.0018
	0.1	.1	−1.429	.012	−.0048
		.2	−1.145	.010	−.0039

It can be observed that if we begin with the Orr and Mellon case 3, and increase M from 0.0 to 0.1 to 1.0, the value of $\hat{D}$ drops from 12.7 to 10.7 to 9.8, and the probability of deficiency from .33 to .01 to .001. A second point is that the sensitivity of $\hat{D}$ to other parameters falls quickly, the larger is M. Finally, the Orr and Mellon conclusion that uncertainty must lower the expansion multiplier is shown to be wrong by the marked (a) counterexamples. With the aid of the second-order conditions for a local maximum, it is easy to see analytically why one may be led numerically to this false conclusion. Rewrite (6) as:

$$\hat{D} < \frac{\dfrac{R}{1-\rho} + \dfrac{r(1-\rho)}{M}}{\left(\dfrac{\rho}{1-\rho} + k\right)}$$

or

$$\text{(11)} \quad \hat{D} < D^* + \frac{r(1-\rho)}{M\left(\dfrac{\rho}{1-\rho} + k\right)}$$

If M is so large that it dominates the values of other parameters, then we have approximately

$$\text{(12)} \qquad \hat{D} < D^*$$

If, on the other hand, small values of M are considered, it is apparent that $\hat{D}$ can exceed D^*.[8]

---

[8] It can be shown further that $\hat{D}<D^*$ if, and only if, $\alpha<.5$. This conclusion is independent of R essentially because α is not a function of R, but of the parameters i, ρ, k, M, and r only. For example, the α columns of Tables 2 and 3 in the text are identical, although R changes from 10.0 to 1.0.

Orr and Mellon were misled by a poor experimental design. It is obvious that one cannot examine a grid of all combinations of possible levels of parameters; nevertheless, it is imperative that key factors be isolated and examined. To demonstrate that the solution for $\hat{D}$ is still inadequately explored, Table 3 is presented. Here the initial excess reserves, R, are set at 1.0, the variance of reserve losses (i is maintained at .0025 and k at .7). Not surprisingly, there are many cases which require a cutback, rather than expansion, to maximize expected profits.

REFERENCES

D. Orr and W. G. Mellon, "Stochastic Reserve Losses and Expansion of Bank Credit," *Amer. Econ. Rev.*, Sept. 1961, *51*, 614–23.

——— and ———, "Errata," *Amer. Econ. Rev.*, Sept. 1963, *53*, 745.

Discrimination and Income Differentials: Comment

By ORLEY ASHENFELTER AND MICHAEL K. TAUSSIG*

In a recent article in this *Review*, James Gwartney states as his basic premise: "When measuring employment discrimination, the relevant comparison is between individuals of similar productive capacity who differ only in color." He then proceeds to decompose gross nonwhite/white income differentials into two components: " . . . (a) a differential resulting from differences in productivity factors not directly related to employment discrimination, and (b) a residual unaccounted for by differences in productivity factors and which may result largely from employment discrimination" (see Gwartney (1970, p. 396)). Beginning from an unadjusted nonwhite/white income ratio of .583 for urban males in 1959, he finds, after successive, cumulative adjustments for differences in the "explanatory factors" of quantity of education, scholastic achievement levels, regional distribution, city size, and age distribution, that the adjusted ratio is between .81 and .87, so that between 45 and 31 percent of the gross nonwhite/white income differential can probably be attributed largely to employment discrimination. Gwartney's most novel finding is that differences in scholastic achievement levels explain a large portion of the nonwhite/white income difference, a result derived by adjusting the quantity of education (years of schooling) using the scholastic achievement test scores of students at grades 6, 9, and 12 in 1965 as given in the *Coleman Report.*

We argue below that, in fact, Gwartney's computations provide little, if any, evidenc on the factors *causing* an income differenti to exist between nonwhites and whites, an therefore that the conclusions Gwartne draws from his calculations may be seriousl misleading as a guide to public policy.

I. The Adjustment for Scholastic Achieveme

Since Gwartney's one adjustment f scholastic achievement is easily the singl most important quantitative factor in na rowing the gap between his adjusted an unadjusted nonwhite/white income diffe entials, it is a crucial determinant of hi estimate of current labor market discrim nation. His major results are summarized i Table 1, which we reproduce here withou change from his article. Gwartney obtai his estimate by combining data from th 1960 *Census of Population* on incom classified by schooling level and race wit the 1965 data from the *Coleman Report* o standardized test scores classified by schoo ing level and race. Since Gwartney is unabl to demonstrate any relationship between th scores of students on the *Coleman Repo* tests and the subsequent market produ tivity of those students, it seems intuitivel clear that some fairly strong assumptio will be necessary in order to use *only thes data* to estimate the effect of differences i scholastic achievement on nonwhite/whit income differences. Unfortunately, Gwartne provides no discussion whatever of th conditions under which his calculations pr vide an unbiased estimator of the effect scholastic achievement differences on th nonwhite/white income differential. Thi probably results from Gwartney's failure t specify a conceptual framework detailin the mechanism by which "productivity" an other factors are alleged to affect earning (income). We proceed, therefore, as follow First, we specify a rough framework fo evaluating Gwartney's estimation procedur Second, we examine explicitly the assump

---

* Associate professor of economics, Princeton University, and associate professor of economics, Rutgers College, Rutgers University, respectively. We are grateful to Daniel S. Hamermesh, Albert Rees, and Finis Welch for helpful comments and to the Ford Foundation for financial support of the urban labor markets program of the Princeton University Industrial Relations Section. The authors are solely responsible, of course, for the views expressed in the paper and for any errors.

TABLE 1—INCOME OF NONWHITES AS A PERCENTAGE OF WHITES, ADJUSTED FOR VARIOUS DETERMINANTS OF INCOME DIFFERENTIALS BETWEEN WHITE AND NONWHITE URBAN MALES FOR THE UNITED STATES IN 1959

	Index of Income Differences		*Marginal Effect of Factor*	
	Laspeyres	Paasche	Laspeyres	Paasche
Unadjusted Income Ratio (Nonwhite/White)	58.3	58.3	—	—
Explanatory Factors				
A. Quantity of Education	67.0	69.9	8.7	11.6
B. Scholastic Achievement	85.1	82.1	18.1	12.2
C. State Distribution	91.0	84.5	5.9	2.4
D. City Size	89.3	83.3	−1.7	−1.2
E. Age Distribution	86.5	80.9	−2.8	−2.4

Source: Gwartney (1970, Table 1, p. 402).

tions necessary for establishing that Gwartney's procedure will give an unbiased estimator of the productivity effect on earnings. Finally, for the sake of completeness, we present the results of the application of Gwartney's technique to data for female workers, to supplement the results for male workers reported in Gwartney's article.

Assume that, in the absence of discrimination, productivity and earnings, Y, would depend positively on: the number of years of schooling, S; any "learning" that enhances market productivity, L; and other productivity-enhancing factors, X, such as on-the-job training or health. Thus,

$$Y = f(S, L, X) \tag{1}$$

Since L is unobservable, we may consider using scholastic achievement test scores, T, as a proxy for it. In general, we expect T to depend positively not only on L, but also on S and on other factors, perhaps cultural or environmental, Z, which affect test scores, but *not* market productivity. Thus, $T = g(S, L, Z)$, so that we may write the relationship which will let us use T as a proxy for L as

$$L = h(S, T, Z) \tag{2}$$

Equation (2) cannot be estimated, of course, because L is unobservable, but (2) can be substituted into (1) to obtain

$$Y = f(S, h(S, T, Z), X) \tag{3}$$

Note that $\partial Y/\partial S = f_1 + f_2 h_1$ and $\partial Y/\partial T = f_2 h_2$. According to (3), if we wanted to know how much of the difference in Y between nonwhites and whites with equal values of S, Z, and X was due to differences in L, we could obtain an estimate of this quantity by multiplying the difference in T for the two groups times an estimated $\partial Y/\partial T$. This is true because $dY = (\partial Y/\partial L)dL = f_2 dL = f_2 h_2 dT \equiv (\partial Y/\partial T)dT$ for constant (i.e., equal) values of S, Z, and X.

Ideally, this approach calls for a longitudinal study of a sample of representative individuals in order to obtain joint observations on lifetime earnings streams, test scores, years of schooling, and measures of X and Z.[1] Gwartney, however, has only the *Coleman Report* data giving average student test scores by grade levels 6, 9, and 12 in 1965; and the 1960 *Census of Population* data giving average 1959 incomes of adults by completed grade level. It is natural, therefore, to inquire as to the conditions under which these data alone can be used to obtain an unbiased estimator of the nonwhite/white earnings differential due to differences in L. The necessary assumptions are:

1) Holding Z, T, and X constant in (3),

---

[1] See the recent paper by W. Lee Hansen, Burton Weisbrod, and William Scanlon which attempts to merge the effects of schooling and scores on the Armed Forces Qualification Test on the earnings of a sample of males who failed that test. Also, see the work of O. D. Duncan (1969).

S must have no effect on earnings, i.e., we require that $\partial Y/\partial S=0$ or $-f_1/f_2=h_1$. This means that the additional L required to replace a unit of S in order to increase earnings by some given amount should be identical to the decline in L implied by an increase in S in order to maintain a given test score. This requires a remarkable accuracy in measuring economic productivity for the scholastic achievement tests in the *Coleman Report*. It implies, for example, that if we observe two groups of white workers with 9 and 12 years of schooling, respectively, but with the same test scores, the two groups should have exactly the same earnings. Since S and T will be positively correlated, if S has an independent effect on Y this will result in an upward bias to the estimator of $\partial Y/\partial T$ and thus an upward bias to the estimator of the effects of differences in L on Y.[2]

2) We must assume that the unobserved Z factors are equal for nonwhites and whites; if cultural or environmental factors favor whites over nonwhites in testing, whites will achieve higher test scores than nonwhites with equal market productivity. A review of the literature by P. Wallace, B. Kissinger, and B. Reynolds strongly indicates that most test instruments are culturally biased against nonwhites and are poor predictors of actual job performance. If the assumption of equal Z for nonwhites and whites does not hold, this gives an upward bias to the estimator of the effect of differences in L on Y.

3) Finally, we must assume that the income differences by educational category in the *Census* data accurately reflect what children who took the *Coleman Report* tests would have earned. But the *Census* income data refer to the highest educational category achieved by an individual. For example, the *Census* income figure for grade 6 refers to individuals who completed grade 5 by 1969 *and had no further schooling*. The *Coleman Report* data, on the other hand, give average test scores for students in a given grade in 1965, most of whom can be presumed to have continued their schooling. Only by chance will the average test score for grade 6 (for white students) be representative of the average test score for those whose *highest* grade is grade 6. On the plausible assumption that dropouts at a given grade have a lower average test score than all students who take the test at that grade, and that this difference declines at higher grade levels, then Gwartney's technique for combining the *Census* and *Coleman Report* data will result in an upward bias to his estimator of $\partial Y/\partial T$ and thus in an upward bias to the estimator of income differences due to differences in L.

Only if assumptions 1), 2), and 3) are maintained, are the available data sufficient to provide an unbiased estimator of the difference in earnings between whites and nonwhites due to differences in L. First, the *Coleman Report* data provide the relationship $T_N=N(S)$ for nonwhites and $T_W=W(S)$ for whites. Likewise, the *Census* provides $Y_W=Y(S)$. Thus the observable relationship $Y_W=Y(W^{-1}(T_W))$ provides an estimate of $\delta Y/\delta T$. Comparing $Y_N/Y(W^{-1}(T_N))=Y_N/Y(W^{-1}(N(S)))$ at each schooling level gives an estimate of what the ratio of nonwhite to white earnings *would be* if L were the same for both groups. Note that the observed relationship in the *Coleman Report* between average test score and grade level for whites must be assumed to be the relationship which would exist for nonwhites in the absence of discrimination. Likewise $Y_W-Y(W^{-1}(N(S)))$ is an estimate of the difference in earnings due to differences in L at each schooling level. It is easy to see that, in general, the failure of any of the above assumptions is likely to

---

[2] Suppose we adopt a linear approximation to (3) and estimate $\partial Y/\partial T$ by the partial regression coefficient b_T in $Y=b_TT+b_SS$. Our estimate of b_T in (3) if we did not hold S constant would be $b_T^*=b_T+b_Sb_{ST}$, where b_S is the estimate of $\partial Y/\partial S$ and b_{ST} is the regression coefficient of S on T. In order to obtain a rough notion of the possible proportionate bias, $(b_T^*-b_T)/b_T$, in Gwartney's procedure we have used the only broad sample of which we are aware. Duncan (1968) provides correlations among S, T, and Y from the *Current Population Survey—National Opinion Research Center* data, where T is measured by Armed Forces Qualification Test scores. From Duncan's results we may calculate $(b_T^*-b_T)/b_T=b_Sb_{ST}/b_T=\beta_S\beta_{ST}/\beta_T=(.26)(.59)/(.16)=.96$, where the β's represent the standardized regression coefficients corresponding to the b's. This calculation suggests a large upward bias on the order of 96 percent. The estimate of the effect of differences in L on Y would thus be biased upward by almost 100 percent.

Table 2—Income Ratio of Nonwhites to Whites Adjusted for Various Determinants of Earnings Differentials Between White and Nonwhite Urban Females for the United States in 1959

	Index of Income Differences		*Marginal Effect of Factor*	
	Laspeyres	Paasche	Laspeyres	Paasche
Unadjusted Income Ratio (Nonwhite/White)	65.9	65.9	—	—
Explanatory Factors				
A. Quantity of Education	81.9	81.7	16.0	15.8
B. Scholastic Achievement	124.9	108.5	43.0	26.8
C. State Distribution	137.7	118.4	12.8	9.9
D. City Size	134.7	115.8	−3.0	−2.6
E. Age Distribution	125.3	108.0	−9.4	−7.8

Source: Gwartney (1969; Table 5, p. 31).

bias the estimator of $\partial Y/\partial T$, and hence $Y_W - Y(W^{-1}(N(S))) = \partial Y/\partial T\,[T_W - T_N]$, upward. Thus, the estimator of employment discrimination as the residual income differential would tend to be biased downward.

It is not possible to estimate directly the magnitude of the bias in Gwartney's procedure without additional data. Table 2, however, shows results for urban women in 1959 which are identical to those reported by Gwartney for urban men in 1959 in Table 1. Although not reported in this *Review*, Table 2 was taken *without modification* directly from his doctoral dissertation. Following Gwartney's own interpretation, these results imply that after adjustment for differences in education, scholastic achievement, and other factors, the median income of nonwhite women is estimated at between 125 and 108 percent of that of white women. This curious finding implies that there is labor market discrimination *in favor of* nonwhite women over white women and that it results in incomes that are from 8 to 25 percent greater for the former.

Needless to say, these calculations for female workers, like the ones in Table 1 for male workers, should not be taken seriously until additional information regarding assumptions 1)–3) becomes available. Moreover, additional adjustments could undoubtedly be made to the calculations for females that would give a nonnegative estimate of the extent of discrimination against nonwhite women. Our basic point, however, is that Gwartney's procedure is based on such restrictive maintained hypotheses that the *only* situation in which one can evaluate the quality (or "precision") of his estimates is the case where they give patently absurd results.

II. A Note on the Conceptual Framework

In the interpretation of his calculations Gwartney postulates, as have others, a dichotomy between the effects on income of racial productivity differences and of the residual factor of labor market discrimination. He explicitly recognizes that he must exclude from his list of productivity factors any variables that are directly related to employment discrimination, but he does not see that this same logic might apply equally to at least some of the factors he includes. Specifically, he implicitly assumes a unicausal relationship flowing from years of schooling to income. As several writers have observed, however, educational attainment may well be in part a result of the existence and expected continuation of labor market discrimination. Black youngsters who expect to face unabated labor market discrimination in the future may curtail their formal education, undertake little investment in other forms of human capital, and make "realistic" downward adjustments in their aspirations as aspects of a rational strategy.[3]

[3] See, for example, the argument by S. Michelson along these lines.

Given the above argument, it follows that estimates of the effect of productivity and discrimination factors on nonwhite/white income differences should be worked out within the context of a carefully specified behavioral model which takes full account of the simultaneity and time lags in decision making. No economist has yet specified and estimated the parameters of such a model, and we do not criticize Gwartney for failing to do so. Other writers have, however, squarely faced the problems involved in their partial equilibrium analyses and have explicitly recognized the limitations of their results.[4] Gwartney however concludes from his calculations that "... unless differences between the two populations in these two factors [of the quantity of education and scholastic achievement] can be reduced substantially, the median income of nonwhite males is unlikely to increase to more than 70 to 80 percent of that of whites, even if *employment discrimination is substantially reduced*" (1970, p. 397, emphasis added). The truth is that no economist is yet in a position to justify such an estimate. If the feedback mechanism from labor market discrimination to educational attainment and scholastic achievement suggested above does hold, then even if Gwartney's conclusion were based on a sound estimating technique, it would still be seriously misleading as a guide for policy decisions which have a time horizon long enough to allow the feedback effect to operate. Government policies that eliminate labor market discrimination may well have the important side effect of encouraging nonwhites to invest as much in human capital as whites.

REFERENCES

J. S. Coleman et al., *Equality of Educational Opportunity*, Washington 1966.

O. D. Duncan, "Ability and Achievement," *Eugenics Quart.*, Mar. 1968, *15*, 1–11.

———, "Inheritance of Poverty or Inheritance of Race," in D. P. Moynihan, ed., *On Understanding Poverty: Perspectives from the Social Sciences*, New York 1969, 85–110.

J. D. Gwartney, "Discrimination and Income Differentials," *Amer. Econ. Rev.*, June 1970, *60*, 396–408.

———, "Income and Occupational Differences Between Whites and Nonwhites," unpublished doctoral dissertation, Univ. of Washington 1969.

W. L. Hansen, B. A. Weisbrod, and W. J. Scanlon, "Schooling and Earnings of Low Achievers," *Amer. Econ. Rev.*, June 1970, *60*, 409–18.

S. Michelson, "Rational Income Decisions of Negroes and Everybody Else," *Ind. Labor Relat. Rev.*, Oct. 1969, *23*, 15–28.

D. M. O'Neill, "The Effect of Discrimination on Earnings: Evidence from Military Test Score Results," *J. Hum. Resources*, fall 1970, *5*, 475–86.

P. Wallace, B. Kissinger, and B. Reynolds, *Testing of Minority Group Applicants for Employment*, Office of Research and Reports, Equal Employment Opportunity Commission, Res. Report 1966–67, Mar. 1966.

F. Welch, "Labor-Market Discrimination: An Interpretation of Income Differences in the Rural South," *J. Polit. Econ.*, June 1967, *75*, 225–40.

---

[4] In contrast to Gwartney's discussion, see the thoughtful treatment of these points in F. Welch and Dave O'Neill.

Discrimination and Income Differentials: Reply

By JAMES GWARTNEY*

The comments of Orley Ashenfelter and Michael Taussig (henceforth referred to as A-T) are helpful in clarifying some of the measurement difficulties encountered when one seeks to quantify the impact of employment discrimination on the relative income of nonwhites. They raise three major issues. First, A-T suggest that the contribution of productivity factors to the explanation of racial income differences is overstated because (a) schooling has a significant impact on earnings that is independent of productive ability, (b) achievement tests including those of the *Coleman Report*, understate nonwhite ability because of cultural and environmental factors favorable to whites, and (c) the achievement differential for students at various grade levels does not accurately reflect the actual differential for those in the labor force in 1959, who completed comparable grade levels but no more. Second, A-T note that when the adjustment technique, used to estimate the impact of productive differences is applied to *annual income* data of urban women, the actual income differential is more than explained by productivity differences. This is cited as indirect evidence that the adjustment procedures overstate the impact of productivity differences. In addition, they raise the possibility that employment discrimination may have substantial feedback effects that contribute to productivity differences according to color. I shall discuss each of these points in order.

I. Is the Impact of Productivity Differences Overstated?

Estimates for the adjusted income ratio were derived both by utilization of an index number method and by regression analysis. Similar estimates were obtained in both cases. Ideally, one would like to compare the incomes of individuals (groups) who have the same quantity of education and achievement, but differ only in color. Such data are not presently available. However, both income and achievement data are available according to color and level of education measured in quantity of years. Therefore, it was possible to estimate the relative income of groups of individuals with similar achievement. However, when the index method is utilized the two groups have a different quantity of education.[1] If quantity of education has a positive impact on income, independent of its contribution through promotion of higher achievement levels, the achievement-adjusted income ratio will be biased upward when estimated by the index method. However, it is not obvious that this is the case.[2] One can argue, as A-T do indirectly, that there is a type of "sheepskin effect" operating, and therefore employers would be expected to prefer employees with more education to those with less even though their achievement levels are identical. But there may also be a countereffect, what I would term the "rapidity of learning effect." If the achievement levels of two workers are equal but their quantity of schooling is different, this suggests a difference in their ability to learn rapidly—a trait

---

* Assistant professor of economics, The Florida State University. I have benefited from discussions with various colleagues including John Chang, Charles Rockwood, and Howard Tuckman.

[1] When the regression method is utilized, adjustment is made simultaneously for quantity of education, achievement differences and other productive factors. Therefore, the problem discussed here does not arise when the estimates are derived by regression analysis.

[2] See W. Lee Hansen, Burton A. Weisbrod, and William J. Scanlon for evidence that education, independent of achievement, has an insignificant impact on earnings, at least in the case of those who failed the Armed Forces Qualification Tests (*AFQT*). As recently as 1965, of 383,000 18-year olds examined, 19 percent of white testees and 67 percent of nonwhites failed the *AFQT*. After adjustment for various productivity factors including education and achievement, the income of nonwhites was found to be 80 percent of the total for this group of males.

which may have an important bearing on worker productivity. *Ceteris paribus*, employers would be expected to prefer the more rapid learners. The impact of these two factors will be in opposite directions, and a priori, it is not obvious which will dominate.

The method outlined by A-T, utilizing partial correlation coefficients, is incapable of estimating the extent of any bias in income comparisons according to color between groups of equal achievement but a different quantity of education. This issue is related to the multicollinearity problem. Since achievement and education are related to each other (partial $r=.59$ in the data used by A-T),[3] one cannot determine the independent impact of either of the two variables.[4] The method of A-T implicitly assumes that all of the change in income associated with additional years of education results because of that education. This method will implicitly assign some of the additional income that should be attributed to achievement, to the educational factor because of the high degree of collinearity between the two. It should be noted that this problem is diminished by the regression method that was utilized in my paper. The achievement variable was the *differential* in achievement between whites and nonwhites, a variable that is not highly correlated with quantity of education.

But even if multicollinearity was not a problem, one cannot use the relationship between quantity of education (measured in years) and achievement for *whites* to argue that the achievement adjustment is upward biased. The major reason for making the adjustment in the first place is that the relationship between achievement and quantity of education is different for blacks because they receive both a lesser quantity (measured in days) and quality of education than whites of a similar grade level.[5]

Second, A-T argue that the achievement correction may overstate the adjusted nonwhite/white income ratio because the achievement tests used may, on balance, favor whites. While determination of whether or not achievement tests such as those utilized by the *Coleman Report* are actually biased against nonwhites is well beyond the objectives of this paper, it should be noted that the market, like the tests, will have certain cultural and environmental biases, independent of color discrimination. The market probably favors those who possess larger vocabularies, utilize the more widely accepted grammar and colloquialism; and who have the characteristics that are held by most employers, consumers, and fellow employees. Such attributes may not directly effect one's marginal revenue product, but because of the preferences of others, lack of such preferred characteristics will reduce one's earning ability regardless of color. The point is not that color discrimination is unimportant. It obviously is. But, rather, that while it may be true that the achievement tests may favor those with the cultural background and attributes of the representative individual, it is also true that the employment market, *independent of color*, will favor the same representative individual.

A third possible source of bias is introduced if the achievement differential for those presently in school in 1965 does not accurately reflect the differential for those in the labor force in 1959. Problems introduced by the imperfect cohort comparison were discussed in the paper. It was pointed out that if the differential has declined over time, due to, say, increases in relative expenditures according to color on educational inputs, the estimated achievement gap, and therefore the achievement-adjusted relative income estimates, would both be biased downward. However, this is not the problem raised by A-T. Rather they question the application of data for individuals presently in school and who are likely to continue, to individuals in the labor force who have completed a given grade, but no more. This factor poses a problem only if there is reason to expect the achievement *differential* for those presently in school is a biased estimate of the *differential* for "dropouts." While one would expect the mean achievement level of dropouts to be lower than for those who continue, this is equally true for both whites and nonwhites. If, as would seem reasonable, the dropouts are drawn in a similar manner from the two achievement distributions,[6] the estimated mean achievement differential for all students would be an unbiased estimate of the mean differential for dropouts only.

---

[3] The data utilized by A-T were different from those of my original paper in two important respects. They were limited to white males 25–34, and performance on the *AFQT*, rather than the data of the *Coleman Report*, was used as the measure of achievement. See Otis D. Duncan.

[4] See Arthur Goldberger, pages 192–94 and 197–200, for an elaboration of this point.

[5] Nonwhites with equal quantity of education measured in years will have fewer days of attendance than whites because of their over-representation in schools (largely in the South) with a six- to eight-month term and a seven-year elementary degree requirement. Also the absentee rate of blacks during the relevant period exceeded that of whites since black children were more often utilized as field hands during harvest periods.

In addition, the work of Dave O'Neill enhances our confidence in the applicability of the Coleman data to those in the labor force. Utilizing data from the *AFQT*, for individuals who have *completed* various grade levels, O'Neill estimated the performance level of blacks to be 2.8 years less than whites at grade 9, 3.5 years at grade 12, and 4.2 years for college graduates. These estimates are remarkably similar to the findings of Coleman. Their similarity suggests they are reliable estimates of achievement differences between whites and nonwhites of various grade levels who completed their schooling during the 1950's and 1960's. While we would indeed prefer longitudinal estimates of the differential, until such estimates are made we must utilize the best available data. Presently, this would appear to be the data of O'Neill and Coleman.

In summary neither the magnitude nor the direction of the bias on the unexplained residual introduced by the three potential sources cited by A-T can be determined. In addition, the consistency of my estimates, utilizing two different methods, with the estimates of others suggest that current productivity factors are responsible for between two-fifths and two-thirds of current income differences.[7]

II. The Adjusted Income Ratio for Females

Ashenfelter-Taussig are correct that when annual income data for females are used the index number technique does yield an adjusted nonwhite/white income ratio greater than unity. But, as explained in my original work quoted by A-T, this is due to differences in the labor market role of white and nonwhite females. The nonwhite female is more likely to be a primary, the white a secondary, income earner. The white female because of her greater family income, will be expected to place more emphasis on nonpecuniary, rather than strictly pecuniary income.[8] As a primary income source, nonwhite females may be more likely to possess job dependability characteristics preferred by employers (low turnover and absentee rate, little opposition to position change within the company, etc.).

The bias introduced by these factors will be reduced if hourly earnings rather than annual income data are used. Utilizing the

---

[6] This merely means that if x percent of the bottom 10 percent (or second, third, etc. decile) for whites dropout, that cx percent of nonwhites in the given decile also drop out. It is not necessary that the relative dropout rates according to color are the same, but only that the c, relative differences, remain constant between the two distributions. Even if c does vary, the direction of the bias will not necessarily be upward. Only if a larger share of the nonwhite dropouts are drawn from the *upper* portion of their achievement distribution will the differential for all students overestimate the differential for dropouts only. If a greater share of the nonwhite dropouts are from the lower portion of this distribution, the bias will be in the opposite direction.

[7] See O'Neill, Hansen, Weisbrod, and Scanlon and Dave Rasmussen for estimates of the nonwhite/white income ratio adjusted for various productivity factors.

[8] If nonpecuniary factors are more important to white females, we would expect them to work both fewer hours per week and weeks per year. The time worked data according to sex in 1960 were:

	Mean hours worked per week	Mean weeks worked per year
White Males	43.5	48.1
Nonwhite Males	39.4	46.7
White Females	35.7	46.5
Nonwhite Females	42.0	46.5

No adjustment for time worked for males was made because this factor may be a function of current employment discrimination. However, in the case of females, time worked differences are likely to result because of the greater preference for leisure on the part of white females who have a higher family income (see Gwartney (1969)).

recent data derived by Victor Fuchs, it is now possible to estimate the adjusted relative earnings according to color using hourly data, pp. 39–46. Using the identical adjustment methods previously applied to the annual income data, Table 1 presents estimates of the adjusted relative hourly earnings for both males and females.

After the data are adjusted for age, education, and achievement, the nonwhite/white hourly earnings ratio is estimated to be .84 for males and between .83 and .87 for females.[9] While utilizing hourly rather than annual earnings data makes little difference for males, it makes a substantial difference for females. Application of the adjustment technique to hourly data suggest that the estimated adjusted nonwhite/white income ratio of greater than unity for females stems, not from bias in the technique, but from differences in the labor market roles—particularly the time worked factor—of nonwhite females relative to whites. The results obtained for females when the technique is applied to hourly earnings data are consistent with what one would expect and enhance our confidence in the method.

TABLE 1—HOURLY EARNINGS OF NONWHITES RELATIVE TO WHITES ADJUSTED FOR AGE, EDUCATION, AND SCHOLASTIC ACHIEVEMENT DIFFERENCES FOR NONAGRICULTURAL EMPLOYED PERSONS IN 1959

	Index of Hourly Earnings Differences (percent)	
	Male	Female
Unadjusted (Nonwhite/White)	63.9	65.8
Explanatory Factors		
A. Age		
Laspeyres	64.3	64.6
Paasche	62.5	65.9
B. Education		
Laspeyres	70.0	77.2
Paasche	71.6	73.3
C. Age-Education		
Laspeyres	72.1	74.4
Paasche	74.2	72.6
D. Age-Education-Achievement		
Laspeyres	84.3	87.9
Paasche	84.2	83.1

Source: Derived from Fuchs.

III. Employment Discrimination and Productivity Differences

As indicated by A-T, one should use caution before drawing policy implications from the nonwhite/white income ratio adjusted for *current* productivity differences. Employment discrimination may have secondary effects that contribute to future productivity differences. *Past* employment discrimination may very well have reduced the incentive of nonwhites to acquire various productive factors and may therefore still exert indirect influence on the *present* economic status of nonwhites. The methods utilized in my paper were not designed and are not capable of measuring the impact of past employment discrimination on present income differences. But the fact still remains that differences in current productivity do contribute substantially to current differences in economic status according to color and unless these differences are reduced, substantial income differences will remain even if the impact of current employment discrimination is minimal.

In addition, one should remember that the estimates are for the nonwhite/white income ratio after adjustment for various current productivity differences. If current employment discrimination diminishes nonwhite income, an unexplained residual would be expected to remain. Such a residual was found, but this residual is not the same thing as employment discrimination. Due to data limitations, it was not possible to adjust for other productivity factors that might either increase or reduce the adjusted differential. The impact of discrimination, even current employment discrimination, may be either greater or less than the unexplained residual.

REFERENCES

J. S. Coleman et al., *Equality of Educational Opportunity*, Washington 1966.

---

[9] See Robert E. Hall for additional evidence that the adjusted *hourly* wage of black females is less than their white counterparts.

O. D. Duncan, "Ability and Achievement," *Eugenics Quart.*, Mar. 1968, *15*, 396–408.

V. R. Fuchs, *Differentials in Hourly Earnings by Region and City Size*, Occas. Paper 101, Nat. Bur. Econ. Res., New York 1967.

A. S. Goldberger, *Econometric Theory*, New York 1964.

J. D. Gwartney, "Income and Occupational Differences between Whites and Nonwhites," unpublished doctoral dissertation, Univ. Wash. 1969.

R. E. Hall, "Why is the Unemployment Rate so High at Full Employment?" in A. M. Okun and G. L. Perry, eds., *Brookings Papers on Economic Activity (3)*, Washington 1970.

W. L. Hansen, B. A. Weisbrod, and W. J. Scanlon, "Schooling and Earnings of Low Achiever," *Amer. Econ. Rev.*, June 1970, *60*, 409–18.

D. M. O'Neill, "The Effect of Discrimination on Earnings: Evidence from Military Test Score Results," *J. Hum. Resources*, fall 1970, *5*, 475–86.

D. Rasmussen, "Discrimination and the Income of Nonwhite Males," *J. Econ. Soc.*, forthcoming.

NOTES

EIGHTY-FOURTH ANNUAL MEETING OF THE AMERICAN ECONOMIC ASSOCIATION

New Orleans, Louisiana, December 26–29, 1971

Preliminary Announcement of the Program

Sunday, December 26, 1971

2:00 P.M. EXECUTIVE COMMITTEE MEETING

Monday, December 27, 1971

8:30 A.M. ON THE STATUS AND RELEVANCE OF ECONOMIC THEORY
Chairman: RICHARD A. LABARGE, Louisiana State University, New Orleans
Papers: JOHN R. MORONEY, Tulane University
The Current State of Money and Production Theory
ANDREW BRIMMER, Board of Governors of the Federal Reserve System
The Current Relevance of Theory to Public Policy
Discussants: ROBERT L. BRITE, Louisiana State University, New Orleans
R. WILLIAM PFOUTS, University of North Carolina, Chapel Hill
(Others to be announced)

8:30 A.M. ON THE EMERGING PROBLEMS OF DEVELOPMENT POLICY
Chairman: HOLLIS B. CHENERY, International Bank for Reconstruction and Development
Papers: ALBERT FISHLOW, University of California, Berkeley
Brazilian Income Distribution
R. ALBERT BERRY, Yale University
Income Distribution and Employment
Discussants: NICHOLAS G. CARTER, International Bank for Reconstruction and Development
(Others to be announced)

8:30 A.M. THE STOCK MARKET AND THE ECONOMY
Chairman: HOWARD STEIN, The Dreyfus Fund
Papers: IRWIN FRIEND, University of Pennsylvania
The Economic Consequences of the Stock Market
ROBERT RASCHE, Federal Reserve Bank of St. Louis
Impact of the Stock Market on Private Demand
Discussants: JAMES S. DUESENBERRY, Harvard University and the Federal Reserve Bank of Boston
JAMES H. LORIE, University of Chicago
DAVID I. MEISELMAN, Virginia Polytechnic Institute and State University
FRANCO MODIGLIANI, Massachusetts Institute of Technology

8:30 A.M. GRADUATE STUDENT PAPERS (Sponsored by *The American Economist* and Omicron Delta Epsilon)
Chairman: JOHN D. GUILFOIL, New York University
Papers: (To be announced)

10:30 A.M. THE ECONOMICS OF FULL RACIAL EQUALITY
Chairman: CHARLES L. SCHULTZE, The Brookings Institution
Papers: ALBERT WOHLSTETTER AND SINCLAIR COLEMAN, The Rand Corporation
Alternative Theories of Discrimination: The Policy Implications
DURAN BELL, The Brookings Institution
Occupational Discrimination as a Source of Income Differences—And What Racial Equality Would Mean
Discussants: (To be announced)

10:30 A.M. ECONOMISTS AND THE TRADE UNION MOVEMENT
Chairman: NAT WEINBERG, United Automobile Workers

Papers: ROBERT LEKACHMAN, State University of New York at Stony Brook
Academic Wisdom and Union Reality
BRENDAN SEXTON, New York University
The Working Class Experience
Discussants: NATHANIEL GOLDFINGER, AFL-CIO
(Others to be announced)

10:30 A.M. SOCIAL INDICES OF ECONOMIC PERFORMANCE (Joint Session with the Association for the Study of the Grants Economy)
Chairman: (To be announced)
Papers: MARTIN PFAFF, Wayne State University and Augsburg University
Measurement of Utility
ANITA B. PFAFF, Wayne State University and Augsburg University
An Index of Consumer Satisfaction
A. COSKUN SAMLI, Virginia Polytechnic Institute
Indicators of Changes
B. STRUMPEL, University of Michigan
Indicators of Interaction Between Man and Economic Environment
Discussants: (To be announced)

1:00 P.M. PARTICIPANTS' LUNCHEON*
Co-Chairmen: JAMES TOBIN, Yale University AND JOHN KENNETH GALBRAITH, Harvard University
Toast to Guest of Honor and First American Nobel Memorial Prize Winner, Paul A. Samuelson:
SIR ERIC ROLL, Warburg's Ltd. and the Bank of England
Response: PAUL A. SAMUELSON, Massachusetts Institute of Technology

* Ticket arrangements and charges to be announced.

2:30 P.M. THE ECONOMICS OF THE MILITARY-INDUSTRIAL COMPLEX
Chairman: MARC J. ROBERTS, Harvard University
Papers: WALTER ADAMS, Michigan State University
The Military-Industrial Complex as a Power Problem
RICHARD KAUFMAN, The Joint Economic Committee of the U.S. Congress
MIRVing the Boondoggle: Contracts, Subsidy, and Welfare in the Aerospace Industry
MICHAEL REICH, Harvard University
Does the U.S. Economy Require Military Spending?
JAMES R. KURTH, Harvard University
The Political Economy of Weapons Procurement

2:30 P.M. LONG-TERM TRENDS: THREE ECONOMIES IN THE YEAR 2000 (Joint Session with the Association for Comparative Economics)
Chairman: EVSEY D. DOMAR, Massachusetts Institute of Technology
Papers: T. S. KHACHATUROV, Academy of Sciences of the U.S.S.R.
Long-Term Trends in the Soviet Economy: The U.S.S.R. in 2000
LESTER C. THUROW, Massachusetts Institute of Technology
Long-Term Trends in the American Economy: The United States in 2000
BENJAMIN WARD, University of California, Berkeley
Long-Term Trends in the Yugoslavian Economy: Yugoslavia in 2000
Discussants: ROBERT L. HEILBRONER, New School for Social Research
PETER MIOVIC, Cornell University
NORMAN M. KAPLAN, University of Rochester

2:30 P.M. INTRA-HOUSEHOLD AND INTRA-ORGANIZATION GRANTS (Joint Session with the Association for the Study of the Grants Economy)
Chairman: (To be announced)
Papers: JAMES N. MORGAN and associates, University of Michigan
Changing Patterns of Intra-Family Transfers
DAVID C. WARNER, New York City Health and Hospital Corporation and Wayne State University
Grant Elements Inside Large Organizations: The Case of the New York City Hospital System
JESSE BURKHEAD, DALE TAUSSING, AND JERRY MINER, Syracuse University
Budgeting as an Intra-Organizational Granting Process
Discussants: (To be announced)

8:00 P.M. RICHARD T. ELY LECTURE
Chairman: JOHN KENNETH GALBRAITH, Harvard University
Speaker: JOAN ROBINSON, University of Cambridge and University of Waterloo

Tuesday, December 28, 1971

8:30 A.M. SOME CONTRADICTIONS OF CAPITALISM
Chairman: PAUL M. SWEEZY, Editor, *Monthly Review*
Papers: HARRY CLEAVER, Stanford University
Alternative Possibilities for the Third World
(Speaker to be announced)
Poverty and the Dual Economy
Discussants: (To be announced)

8:30 A.M. INEQUALITY: THE PRESENT TENDENCY AND THE REMEDY
Chairman: MARY JEAN BOWMAN, University of Chicago
Papers: T. PAUL SCHULTZ, The Rand Corporation
Long-Term Changes in Personal Income Distribution: Mythology, Fact, and the Explanations
HAROLD M. HOCHMAN, The Urban Institute
Adjustments in Income Distribution: The Role of Public Choice
Discussants: JACOB MINCER, Columbia University
STEPHAN MICHELSON, Harvard University
JOSEPH E. STIGLITZ, Yale University

8:30 A.M. INFLATION IN UNDEREMPLOYMENT: THE INSTITUTIONAL FRAMEWORK (Joint Session with the Association for the Study of the Grants Economy)
Chairman: KENNETH E. BOULDING, University of Colorado
Papers: JANOS HORVATH, Columbia University and Butler University
Institutional Inducement to Inflation: The Role of Misapplied Grants
ALFRED S. EICHNER, Columbia University
Monopolistic Practices and Inflation
STEFAN H. ROBOCK, Columbia University
Inflation as a Global Experience
Discussants: ROBERT J. GORDON, University of Chicago
LASZLO L. ECKER-RACZ, U.S. Treasury Department
RAYMOND F. MIKESELL, University of Oregon
PAUL JONAS, University of New Mexico
GORDON TULLOCK, Virginia Polytechnic Institute

10:30 A.M. THE ARTS IN THE AFFLUENT SOCIETY
Chairman: ABBA P. LERNER, Queens College, City University of New York
Papers: TIBOR SCITOVSKY, Stanford University
The Arts and the Society: What is Wrong
HY FAINE, University of California, Los Angeles
Unions and the Arts
Discussants: IRVING CHESKIN, League of New York Theaters
THOMAS G. MOORE, Michigan State University
WILLIAM J. BAUMOL, Princeton University and New York University

10:30 A.M. WHAT ECONOMIC EQUALITY FOR WOMEN REQUIRES
Chairman: BARBARA R. BERGMANN, University of Maryland
Papers: FRANCINE B. WEISSKOFF, Trinity College and Harvard University, AND HARRIET ZELLNER, Columbia University
"Women's Place" in the Labor Market
CHARLOTTE D. PHELPS, Temple University
Is the Household Obsolete?
Discussants: (To be announced)

10:30 A.M. ECONOMISTS CONSIDER ECONOMIC REPORTERS AND VICE VERSA: A DISCUSSION
Co-Chairmen: WALTER W. HELLER, University of Minnesota AND LEONARD S. SILK, *The New York Times*
Participants: HENDRIK S. HOUTHAKKER, Council of Economic Advisors and Harvard University
JOHN R. MEYER, National Bureau of Economic Research

JOHN A. SCHNITTKER, Robert R. Nathan Associates
EILEEN SHANAHAN, *The New York Times*
BERNARD NOSSITER, *The Washington Post*
RICHARD JANSSEN, *The Wall Street Journal*

1:00 P.M. LUNCHEON IN HONOR OF GUNNAR AND ALVA MYRDAL*
Chairman: EDGAR B. STERN, JR., New Orleans
Response to introduction: GUNNAR MYRDAL

* Ticket arrangements and charges to be announced.

2:30 P.M. HAVE FISCAL AND/OR MONETARY POLICIES FAILED?
Chairman: ROBERT EISNER, Northwestern University
Papers: MILTON FRIEDMAN, University of Chicago and National Bureau of Economic Research
JOHN G. GURLEY, Stanford University
ARTHUR M. OKUN, The Brookings Institution
Discussants: MARTIN J. BAILEY, University of Rochester and The Brookings Institution
JOHN R. BUNTING, The First Pennsylvania Bank in Philadelphia
ROBERT M. CLOWER, University of California, Los Angeles

2:30 P.M. ISSUES AND PERSPECTIVES OF BLACK POLITICAL ECONOMY
Chairman: CHARLES Z. WILSON; University of California, Los Angeles
Papers: MATTHEW HOLDEN, JR., University of Wisconsin, Madison
Political and Economic Change in the Black Community
ROBERT S. BROWNE, Black Economic Research Center
The Economic Case for Reparations to Black America
Discussants: BENNETT HARRISON, University of Maryland
(Others to be announced)

2:30 P.M. THE KEYNESIAN REVOLUTION AND ITS PIONEERS
Chairman: WALTER S. SALANT, The Brookings Institution
Papers: ALAN R. SWEEZY, California Institute of Technology
The Keynesians and Government Policy, 1933–1939
BYRD L. JONES, University of Massachusetts
The Role of Keynesians in Wartime Policy and Postwar Planning, 1940–1946
Discussants: LAUCHLIN B. CURRIE, Department of National Planning, Colombia
MARRINER S. ECCLES, First Security Corporation, Salt Lake City
RICHARD V. GILBERT, Harvard University
ALVIN H. HANSEN, Professor Emeritus, Harvard University
SEYMOUR E. HARRIS, University of California, San Diego
LEON KEYSERLING, Washington, D.C.
ROBERT R. NATHAN, Robert R. Nathan Associates

8:00 P.M. PRESIDENTIAL ADDRESS
Chairman: SEYMOUR E. HARRIS, University of California, San Diego
Speaker: JAMES TOBIN, Yale University

9:15 P.M. BUSINESS MEETING

Wednesday, December 29, 1971

8:30 A.M. TAXATION OF THE POOR AND THE RICH
Chairman: JOSEPH A. PECHMAN, The Brookings Institution
Papers: DAVID M. GORDON, National Bureau of Economic Research
Taxation of the Poor
EARL A. THOMPSON, University of California, Los Angeles
Taxation of Wealth and of the Wealthy
Discussants: ARNOLD C. HARBERGER, University of Chicago
ROBERT J. LAMPMAN, University of Wisconsin
(Others to be announced)

8:30 A.M. IS THE SOUTH STILL BACKWARD?
Chairman: VIVIAN W. HENDERSON, Clark College

Papers: M. I. Foster, Tennessee Valley Authority
Is the South Still a Backward Region, and Why?
F. Ray Marshall, University of Texas
Causes and Remedies in Southern Economic Development
Discussants: Irvin Sobel, Florida State University
James A. Hefner, Clark College and Princeton University

8:30 A.M. THE FUTURE OF CONSUMER SOVEREIGNTY
Chairman: Robert Solow, Massachusetts Institute of Technology
Papers: Abba P. Lerner, Queens College, City University of New York
(Title to be announced)
Herbert Gintis, Harvard University
Commodity Fetishism and Irrational Production: The Framework for an Analysis of Consumer Behavior
Discussants: (To be announced)

10:30 A.M. HISTORY OF ECONOMIC THOUGHT
Chairman: William D. Grampp, University of Illinois at Chicago Circle
Papers: Craufurd D. Goodwin, Duke University
Economic Theory and Society: A Plea for Process Analysis
William R. Allen, University of California, Los Angeles
Irving Fisher, FDR, and the Great Depression
Harry G. Johnson, University of Chicago
The Early Economics of Keynes

10:30 A.M. ECONOMIC GROWTH AND RURAL POVERTY (Joint Session with the American Agricultural Economics Association)
Chairman: T. T. Williams, Southern University, Baton Rouge
Papers: Carl Gotsch, Harvard University
Technical Change and the Distribution of Income in Rural Areas
Emiel Owens, Prairie View A. & M.
The Poverty Programs of the 1960's: An Evaluation
Discussants: Joesph D. Coffey, U.S. Department of Agriculture
J. Martin Redfern, University of Arkansas

10:30 A.M. MODELS FOR URBAN LAND USE, HOUSING, AND TRANSPORTATION
Chairman: Leon N. Moses, Northwestern University
Papers: Jerome Rothenberg and John R. Harris, Massachusetts Institute of Technology
A Simulation Model of Intra-Metropolitan Location: Housing, Business, Transportation, and Local Government
Leonard Merewitz, University of California, Berkeley
Public Transportation: Wish-Fulfillment and Reality in the San Francisco Bay Area
Discussants: Richard F. Muth, Stanford University
David F. Bradford, Princeton University
Charles D. Baker, U.S. Department of Transportation
Benjamin Chinitz, Brown University

2:30 P.M. THE REGULATION OF TRANSPORATION (Joint Session with the Transportation and Public Utility Group)
Chairman: (To be announced)
Papers: George W. Hilton, University of California, Los Angeles
What Regulatory Commissions Do and What They Think They Do
Roy J. Sampson, University of Oregon
Inherent Advantages Under Regulation
Discussants: Grant M. Davis, Auburn University
John C. Spychalski, Pennsylvania State University

2:30 P.M. JOHN R. COMMONS SESSION AND OMICRON DELTA EPSILON BIENNIAL CONVENTION
Chairman: Ervin K. Zingler, University of Houston
Speaker: Milton Friedman, University of Chicago and National Bureau of Economic Research
(Papers and discussants to be announced)

2:30 P.M. ECONOMIC EDUCATION (Committee on Economic Education of the American Economic Association)
Chairman: GEORGE LELAND BACH, Stanford University
Papers: ALLEN C. KELLEY, University of Wisconsin
T.I.P.S. and Technical Change in Classroom Instruction
RICHARD ATTIYEH, University of California, San Diego AND KEITH LUMSDEN, Stanford University
Some Modern Myths in Teaching Economics: The U.K. Experience
MICHAEL ZWEIG, State University of New York at Stony Brook
A New Course in Radical Economics
Discussants: JAMES D. GWARTNEY, Florida State University
JAMES M. BUCHANAN, Virginia Polytechnic Institute
GORDON K. DOUGLASS, Pomona College

2:30 P.M. THE ECONOMICS OF TECHNICAL CHANGE (provisional)
(Papers and discussants to be announced)

Program Notes

This year sessions of the American Economic Association—as distinct from joint sessions originating with sister organizations—will be limited, with a few exceptions, to two main papers and discussions. Discussion papers, the session on Keynes apart, will not normally be published. This change allows a larger number of topics with wider coverage of member interest. As before, main papers will not exceed 4000 words. Where there are more than two papers, participants are being asked to limit their paper to 3000 words.

Note that certain joint sessions are still to be announced.

The attention of members is directed to two luncheons: one honoring the first American Nobel Memorial Prize winner, Professor Paul Samuelson, and Mrs. Samuelson; and the other honoring Gunnar and Alva Myrdal who are returning after a third of a century as guests of the Association to the scene of Professor Myrdal's work on *The American Dilemma.*

The Association has been asked, depending on expressed need, to establish a day care center for children. Would members who have specific intention of using such a center, please communicate with J. K. Galbraith, 207 Littauer Center, Harvard University, Cambridge, Massachusetts 02138.

ANNOUNCEMENTS

Editor's Announcement

Beginning January 1, 1972, the *American Economic Review* will require a submission fee of $10 to accompany each manuscript when it is first sent in for publication, regardless of size. However, manuscripts should not exceed fifty typewritten pages of text. Checks or money orders should be made payable to the American Economic Association.

The submission fee was approved by the Executive Committee of the AEA at its March 1971 business meeting. The fee has two purposes: First, it will help defray costs of handling and screening the submitted papers. Second, it will encourage authors to submit their work in more finished form and hopefully cut down on rejections which are based on the unfinished or preliminary quality of the papers.

Nominating Committee of The American Economic Association

In accordance with Section IV, paragraph 2, of the bylaws of the American Economic Association as amended last year, President-elect John Kenneth Galbraith has appointed a Nominating Committee for 1972 consisting of Charles L. Schultze, Chairman, Robert Eisner, William J. Fellner, Barbara Bergmenn, Thaddeus Spratlen, and Robert Lekachman. Attention of members is called to the part of the bylaw reading, "In addition to appointees chosen by the President-elect, the Committee shall include any other member of the Association nominated by petition including signatures and addresses of not less than 2 percent of the members of the Association, delivered to the Secretary before December 1. No member of the Association may validly petition for more than one nominee for the Committee. The names of the Committee shall be announced to the membership immediately following its appointment and the membership invited to suggest nominees for the various offices to the Committee."

Visiting Scientist Program in Economics

The American Economic Association has received a further grant from the National Science Foundation to continue its *Visiting Scientist Program in Economics* during the 1971–72 academic year. The purpose of this program is to stimulate improved teaching and interest in modern economics at colleges and universities whose major focus is at the undergraduate level. Visits to about forty campuses were provided by the program in 1970–71.

The NSF grant provides financing for visits to campuses by distinguished economists who spend a day or so on each visit. The program is planned jointly between the visitor and host institution; customarily it includes a talk to undergraduate major students, perhaps an informal seminar with faculty and students, discussion of current course and research interests of the host faculty, and other activities as may be desirable. While the ultimate concern is to encourage greater interest in, and better teaching of, economics at the undergraduate level, visitor's talks may be on research, current developments in economics, policy issues, or other topics that may seem appropriate to the needs of the particular institution.

This program is under the general direction of the Association's Committee on Economic Education. Professor Phillip Saunders administers the program under the policies established by the Committee. He works with an informal roster of economists around the country who may be available for such visits, depending on their own schedules and the degree of common interest between them and the inviting institutions. Institutions are free to suggest visitors they would especially like to have, so long as these are within a radius of 200 miles or so of the campus, in order to minimize travel costs. The host institution is expected to provide lodging and meals for the visitor when he is on campus; other costs will be covered by the grant.

Schools interested in applications should write to: Professor Phillip Saunders, Department of Economics, Indiana University, Bloomington, Indiana 47401.

The Commission on College Physics announces the publication of the *Proceedings of the Conference on Computers in Undergraduate Science Education* held in Chicago, August 17–21, 1970. These *Proceedings* include papers on the following topics: I. Computational Mode; II. Interactive Computer Graphics; III. Simulation Mode; IV. Analog Computing; V. Computer-Supervised Instruction; VI. Politics and Practice; VII. The Future Appendices.

They are available on request to qualified individuals. When ordering your copy, please specify your discipline, institutional affiliation, and the nature of your interest in the educational use of computers. Requests may be addressed to: American Institute of Physics, State University of New York, Stony Brook, New York 11790.

New Journal

Under the editorship of Benjamin B. Wolman, Long Island University, the *International Journal of Group Tensions* will publish objective studies of racial, religious, ethnic, socioeconomic, and political conflicts. No research method and no methodological approach will be excluded and the *Journal* will publish surveys, experimental and statistical research, historical, sociological, and anthropological studies, as well as clinical, psychoanalytical, and sociopsychological studies of groups and individuals.

In preparation of manuscript the rules of the American Psychological Association journals must be followed. Manuscripts should be submitted to Professor

Florence Denmark, Associate Editor, Hunter College, Psychology Department, 695 Park Avenue, New York, New York 10021. Correspondence concerning subscriptions and other business matters should be addressed to Dr. Clifton Kew, Business Manager, 177 East 77th Street, New York, New York 10021

An International Symposium on Modelling Techniques in Water Resources Systems will be held in Ottawa, Ontario, Canada, on May 9–12, 1972. The Symposium will bring together international experts from all disciplines involved with the development and application of mathematical modelling techniques to the planning, operation, and management of water resources systems. Approximately 70 percent of the papers will be by invitation only, the rest will be selected from unsolicited papers. Some of the suggested fields are: economic and social, ecologic, water quality, estuarine, hydraulic, hydrologic, political, and institutional. The number of participants will be limited and participation to the Symposium will be by invitation only. Further details can be obtained from: Dr. Asit K. Biswas, chairman, Organizing Committee, International Symposium on Modelling Techniques in Water Resources Systems, Water and Renewable Resources Sector, Department of Fisheries and Forestry, Ottawa, Ontario, Canada.

Omicron Delta Epsilon, the Honor Society in Economics, announces the 1971 winners of the Frank W. Taussig and Irving Fisher Awards. The Taussig Award was won by Charles C. Brown of Boston College with his entry, "Income Guarantees and the Incentive Question." The Fisher Award was won by Michael Szenberg of Long Island University with his entry, "The Economics of the Israeli Diamond Industry," which was his Ph.D. dissertation at the City University of New York.

The entries for the fourth year of the competitions should be submitted to the Departmental Selection Committees by January 1, 1972. Finalists will be judged by a Final Selection Board consisting of Professors Maurice Allais, William Baumol, Leonid Hurwicz, Wassily Leontieff, and Egon Neuberger (editor). The Taussig Award consists of $100 and publication in *The American Economist*. The Fisher Award consists of $1,000 and publication of the manuscript. For more information write to: Egon Neuberger, Editor, Department of Economics, State University of New York, Stony Brook, New York 11790.

Ralph H. Oakes of 1500 Chicago Avenue, Apt. 704, Evanston, Illinois 60201 would like to sell a complete set of the *American Economic Review*, Mar. 1956 to date. Interested buyers should send bids to Mr. Oakes at the above address.

Errata

On page 125 of the June 1971 issue we failed to note that the Stillwater Oklahoma meeting on computer science and statistics will take place November 1 and 2, 1971.

Deaths

Bennett A. Dominick, Jr., professor, department of agricultural economics, Cornell University, Apr. 14, 1971.

Robert Fairbanks, associate professor of economics and public administration, State University of New York at Albany, spring 1971.

Michael Heilperin, visiting professor of economics, California State College, Los Angeles, Apr. 7, 1971.

Gad P. Scoville, professor emeritus, department of agricultural economics, Cornell University, Mar. 9, 1971.

Walter D. Smith, assistant dean, Graduate School of Business, Columbia University, Mar. 8, 1971.

Retirements

M. D. Anderson, professor of economics, University of Florida, June 15, 1971.

Goetz A. Briefs, professorial lecturer, department of economics, Georgetown University, May 1971.

Carroll L. Christenson, professor of economics, Indiana University, June 1970.

Morris A. Copeland, professor of economics, State University of New York at Albany.

Joseph Dorfman, professor of economics, Columbia University, June 1971.

Christopher J. Fagan, associate professor of economics, University of Notre Dame, June 1971.

Mark J. Fitzgerald, professor of economics, University of Notre Dame, June 1971.

George N. Halm, professor emeritus, Fletcher School of Law and Diplomacy, Tufts University.

Seymour E. Harris, University of California, San Diego, June 1971.

Lloyd A. Helms, professor of economics, Bowling Green State University, June 1971.

Karl Hobson, Washington State University.

George B. Hurff, professor of economics, University of Florida, June 15, 1971.

Wilfred S. Lake, professor of economics, Northeastern University, June 1971.

Henry S. Miller, professor of economics, Queens College of the City University of New York.

Charles Schertenleib, department of economics, Georgetown University, May 1971.

Louis Shere, professor of economics, Indiana University, June 1970.

James M. Waller, visiting professor of economics, University of Virginia, June 30, 1971.

Simon N. Whitney, professor of economics, New York University, June 1971.

Visiting Foreign Scholars

Pranab K. Bardhan, Indian Statistical Institute, New Delhi: visiting professor of economics, Massachusetts Institute of Technology.

Douglas Fisher, University of Essex: visiting professor of economics, Pomona College, 1971–72.

Francesco Foret, University of Torino: visiting professor of economics, Virginia Polytechnic Institute and State University, fall quarter 1971.

S. Herbert Frankel, Oxford University: visiting professor of economics, University of Virginia, spring semester 1972.

Rolf Funck, Karlsruhe University, West Germany: visiting scholar, department of economics, State University of New York at Binghamton, fall 1971.

Augusto Graziani, Istituto di Economia, Universita di Napoli: visiting professor of economics, University of Michigan, fall 1971.

Ronald M. Hartwell, Nuffield College, Oxford: professor of economics, University of Virginia, fall semester 1971.

Don B. Hemmings, University of Sheffield: lecturer in finance, Graduate School of Business Administration, New York University.

Albert G. Hines, University of Durham, England: visiting professor of economics, Massachusetts Institute of Technology.

Allan G. King, University of Cambridge: lecturer of economics, University of Texas at Austin.

Jing-Min Lin, Nanyang University, Singapore: visiting professor, State University of New York at Albany, 1971–72.

Antal Majthay, Hungarian Academy of Science, Budapest: visiting professor of economics, Purdue University, spring 1971.

Gordon S. Mills, University of Bristol: visiting professor of economics, University of Virginia, 1971–72.

Chikashi Moriguchi, Kyoto Institute of Economic Research, Japan: visiting associate professor, department of economics, University of Pennsylvania, 1971–72.

Admantios Pepelasis, Athens, Greece: visiting professor of economics, Virginia Polytechnic Institute and State University, spring quarter 1971.

Harry Richardson, University of Kent at Canterbury: visiting professor of economics, University of Pittsburgh, Sept.–Dec. 1971.

Michael E. Scorgie, Monash University, Melbourne: visiting professor, department of accounting, School of Business Administration, San Diego State College, 1970–71.

Promotions

John Q. Adams III: associate professor of economics, University of Maryland.

John M. Albertine: assistant professor, department of economics, Mary Washington College, University of Virginia.

Marcus Alexis: professor, department of economics, School of Business Administration, University of California, Berkeley, July 1, 1971.

David J. Allee: professor, department of agricultural economics, Cornell University, July 1, 1971.

Thomas J. Atchison: professor of management, School of Business Administration, San Diego State College, Sept. 1971.

Douglas A. L. Auld: associate professor, department of economics, University of Guelph.

Ralph E. Beals: professor of economics, Amherst College.

Barbara R. Bergmann: professor of economics, University of Maryland.

Melvin K. Bers: professor of economics, State University of New York at Albany.

Phillip G. Bibo: assistant professor, department of conomics, Indiana State University.

Russell L. Block: associate professor of business law, School of Business Administration, San Diego State College, Sept. 1971.

M. E. Bond: associate professor of economics, Arizona State University.

Miltiades Chacholiades: professor of economics, Georgia State University, Sept. 1971.

Gene L. Chapin: associate professor of economics, Ohio University, Sept. 1, 1971.

Christopher K. Claque: associate professor of economics, University of Maryland.

James L. Cochrane: associate professor, department of economics, University of South Carolina.

Craig G. Coelen: assistant professor of economics, Northeastern University.

John G. Cragg: professor, department of economics, University of British Columbia.

Ward S. Curran: professor of economics, Trinity College, Sept. 1, 1971.

Douglas C. Dacy: professor of economics, University of Texas at Austin.

John S. Evans: assistant professor, department of economics, University of Alabama, Feb. 1, 1971.

Laurence P. Feldman: associate professor of marketing, University of Illinois at Chicago Circle.

Robert A. Flammang: professor of economics, Louisiana State University.

Olan D. Forker: professor, department of agricultural economics, Cornell University, July 1, 1971.

Richard F. Fryman: associate professor, department of economics, Southern Illinois University.

Paul Grimaldi: assistant professor of economics, Seton Hall University, Sept. 1971.

Jared E. Hazelton: associate professor of economics, University of Texas at Austin.

James J. Heckman: assistant professor of economics, Columbia University.

John F. Helliwell: professor, department of economics, University of British Columbia.

Dale Hiestand: professor, Graduate School of Business, Columbia University, July 1, 1971.

Raymond Horton: assistant professor, Graduate School of Business, Columbia University, July 1, 1971.

William Hosek: associate professor of economics, Whittemore School of Business and Economics, University of New Hampshire.

Marvin R. Jackson, Jr.: associate professor of economics, Arizona State University.

Robert J. Kalter: associate professor, department of agricultural economics, Cornell University, July 1, 1971.

Samuel C. Kelley, Jr.: professor of economics, Ohio State University.

B. F. Kiker: professor, department of economics, University of South Carolina.

K. Klawe: associate professor of economics, University of Alberta.

Robert L. Knox: professor of economics, Arizona State University.

Ben Korda: professor of economics, University of Alberta.

William E. Laird: professor of economics, Florida State University.

Harvey Lapan: assistant professor of economics, Northeastern University.

John Ledyard: associate professor of economics, Northwestern University.

Axel Leijonhufvud: professor of economics, University of California, Los Angeles.

David Lewing: assistant professor, Graduate School of Business, Columbia University, July 1, 1971.

Kevin M. Lightner: associate professor of accounting, School of Business Administration, San Diego State College, Sept. 1971.

Allyn O. Lockner: professor, department of economics, South Dakota State University, July 1, 1971.

C. A. Knox Lovell: associate professor of economics, University of North Carolina at Chapel Hill.

F. William McElroy: associate professor, department of economics, Georgetown University, Apr. 1971.

Martin C. McGuire: professor of economics, University of Maryland.

H. Michael Mann: professor of economics, Boston College.

Phillip A. May: professor of economics, Northern Michigan University.

Gary K. Michael: associate professor of economics and business administration, University of Vermont.

Dale Mortensen: associate professor of economics, Northwestern University.

Brian M. Neuberger, associate professor of finance, School of Business Administration, San Diego State College, Sept. 1971.

Duane F. Neuman: extension associate professor of economics, North Carolina State University.

M. S. Noorzoy: associate professor of economics, University of Alberta.

Daniel I. Padberg: professor, department of agricultural economics, Cornell University, July 1, 1971.

Peter H. Pearse: professor, department of economics, University of British Columbia.

Martin M. Perline: professor of economics, Wichita State University, Sept. 1971.

William P. Pinna: assistant professor of economics, North Carolina State University.

Marion L. Piotowski: professor, department of finance, College of Business Administration, Texas A&M University, Sept. 1, 1971.

Saul Pleeter: assistant professor of economics, Indiana University.

Fred B. Renwick: professor of finance, Graduate School of Business Administration, New York University.

William A. Reynolds: professor of economics, Windham College, Sept. 1, 1971.

H. Kirk Roberts: assistant professor of economics, Indiana University.

James N. Rosse: associate professor of economics, Stanford University.

Milton R. Russell: professor, department of economics, Southern Illinois University.

Gian S. Sahota: professor of economics, Vanderbilt University.

Lars G. Sandberg: professor of economics, Ohio State University.

Beverly K. Schaffer: associate professor of economics, Emory University, Sept. 1, 1971.

Mary J. Schlinger: associate professor of marketing, University of Illinois at Chicago Circle.

Barney K. Schwalberg: associate professor, department of economics, Brandeis University.

Gerald W. Scully: associate professor, department of economics, Southern Illinois University.

A. Ross Shepherd: professor, department of economics, University of Missouri at Kansas City, Sept. 1971.

William G. Shepherd: professor of economics, University of Michigan.

Heidimarie Sherman: assistant professor of economics, Whittemore School of Business and Economics, University of New Hampshire.

Daniel G. Sisler: professor, department of agricultural economics, Cornell University, July 1, 1971.

Samuel L. Skogstad: professor of economics, Georgia State University, Sept. 1971.

Richard K. Smith: assistant professor, department of economics, Marietta College.

Robert F. Smith: professor of economics, Louisiana State University.

Walter A. Smith: associate professor of economics, College of Business Administration, University of Tulsa.

Frank P. Stafford: associate professor of economics, University of Michigan.

Richard A. Stanford: assistant professor of economics, Furman University.

Robert N. Stearns: assistant professor, department of economics, Connecticut College.

Marvin J. Sternberg: associate professor of economics, State University of New York at Albany.

Bernt P. Stigum: professor of economics, Northwestern University.

J. D. Stolen: associate professor of economics, University of Nebraska at Omaha.

Paul S. Stone: extension associate professor of economics, North Carolina State University.

Howard R. Swaine: professor of economics, Northern Michigan University.

Steven M. Swanson: assistant professor of economics, Northeastern University.

William O. Thweatt: professor of economics, Vanderbilt University.

Arthur B. Treadway: associate professor of economics, Northwestern University.

George Tzannetakis: professor of economics, School of Business Administration, Seton Hall University, Sept. 1971.

Russell S. Uhler: associate professor, department of economics, University of British Columbia.

J. S. Uppal: professor of economics, State University of New York at Albany.

James W. Walker: associate professor of management, School of Business Administration, San Diego College, Sept. 1971.

Janet West: assistant professor of economics, University of Nebraska at Omaha.

Jerome D. Wiest: professor of management science, Rice University.

Heiko deB. Wijnholds: professor of finance, School of Business Administration, San Diego State College, Sept. 1971.

James A. Wilde: associate professor of economics, University of North Carolina at Chapel Hill.

John Wile: assistant professor of economics, Rensselaer Polytechnic Institute, May 1971.

Bruce Wilkinson: professor of economics, University of Alberta.

Richard D. Winkelman: associate professor of economics, Arizona State University.

Klaus H. Wolff: professor of economics, Middlebury College, Sept. 1971.

Robert G. Wyckham: associate professor, department of economics and commerce, Simon Fraser University.

Randolph O. Yeager: professor of economics, Western Kentucky University.

Bruce Wilkinson: professor of economics, University of Alberta.

Administrative Appointments

Glen W. Atkinson: associate professor, chairman, department of economics, College of Business Administration, University of Nevada, July 1970.

James C. Baker: chairman, department of finance, Kent State University, Sept. 1971.

William C. Bonifield: chairman, department of economics, Wabash College, Sept. 1971.

Richard V. Cotter: associate dean of graduate studies, College of Business Administration, University of Nevada, July 1970.

John V. Craven: professor, chairman, department of economics, Middlebury College, July 1971.

Henry L. Custer: assistant professor, chairman, department of accounting, College of Business Administration, University of Nevada, July 1970.

Melvin A. Eggers: chancellor and president, Syracuse University, June 1971.

Kenneth G. Elzinga: assistant dean, College of Arts and Sciences, University of Virginia, Sept. 1971.

Joseph R. Fahey: dean, College of the Holy Cross.

Daniel O. Fletcher: chairman, department of economics, Denison University.

A. Myrick Freeman III: chairman, department of economics, Bowdoin College, spring 1971.

Bert T. Glaze: chairman, department of economics, Marietta College.

Marshall I. Goldman: chairman, department of economics, Wellesley College.

H. Scott Gordon: chairman, department of economics, Indiana University.

Paul T. Hendershot: director of graduate studies in economics and business administration, University of Mississippi.

Patric H. Hendershott: chairman, economics policy committee, Purdue University.

Warren S. Hunsberger: director, Center for Asian Studies, The American University.

Peter B. Kenen: director, international finance section, Princeton University, July 1, 1971.

Dick Kosobud: head, department of economics, University of Illinois at Chicago Circle.

Robert L. Knox: chairman, department of economics, Arizona State University.

Fred H. Leonard: chairman, department of economics, Smith College.

Cliff Lloyd: chairman, department of economics, University of Iowa.

Jack McCullick: chairman, department of economics, Fort Hays Kansas State College, Feb. 1, 1971.

Francis McLaughlin: assistant chairman, department of economics, Boston College, June 1, 1971.

Joseph R. Mason: chairman, department of economics, State University of New York at Brockport.

Hans C. Palmer: acting chairman, department of economics, Pomona College, 1971–72.

Harold Petersen: chairman, department of economics, Boston College, June 1, 1971.

Walter J. Primeaux, Jr.: director, Bureau of Business and Economics Research, University of Mississippi.

Albert Rees: chairman, department of economics, Princeton University, July 1, 1071.

John W. Rowe, Jr.: chairman, department of economics, University of Houston.

Robert Saunders: chairman, department of economics, Kent State University, Sept. 1971.

Robert C. Stuart: acting chairman, department of economics, Douglass College, Rutgers University, 1971–72.

Daniel B. Suits: chairman, board of studies in economics, University of California, Santa Cruz.

Wesley B. Sundquist, U.S. Department of Agriculture: head, department of agricultural and applied economics, University of Minnesota.

Ingo Walter: associate dean, Graduate School of Business Administration, New York University.

Oliver E. Williamson: chairman, department of economics, University of Pennsylvania.

Ben M. Wofford: associate dean of undergraduate studies, College of Business Administration, University of Nevada, July 1970.

Edward A. Zane: professor, chairman, managerial sciences, College of Business Administration, University of Nevada, July 1970.

Appointments

Charles L. Alfrod III: assistant professor of business administration, Furman University.

William R. Allen, University of California, Los Angeles: professor of economics, Texas A&M University.

Akeel H. Alsadi, University of Southern California:

visiting lecturer, department of economics, San Diego State College.

Wayne R. Archer, Indiana University: assistant professor of real estate and urban land studies, University of Florida, Sept. 1971.

G. Christopher Archibald, University of Essex: professor, department of economics, University of British Columbia.

Fred D. Arditti, University of California: associate professor of finance, University of Florida, Sept. 1971.

Michael Balch, University of Chicago: associate professor, department of economics, University of Iowa.

Stephen E. Baldwin, Chico State College: special reports group, Bureau of Labor Statistics, Washington.

William J. Barger: instructor of economics, University of Southern California.

Christopher C. Barnekov, Jr., University of Chicago: instructor, department of economics, Ohio State University.

Peter Barth, Ohio State University: associate professor, department of economics, University of New Mexico.

George Bateman, University of Chicago: assistant professor, department of economics, Marietta College.

Joseph R. Bisignano: lecturer, department of economics, Rutgers-The State University, July 1971.

Stanley W. Black: associate professor of economics, Vanderbilt University, Sept. 1, 1971.

Roger A. Blau: associate professor of business administration, University of North Carolina at Chapel Hill.

Harry Bloch: assistant professor, department of economics, University of British Columbia.

Barry Blueston, Michigan University: instructor, department of economics, Boston College.

Steven Bolten, University of Houston: visiting assistant professor of finance, Graduate School of Business Administration, New York University.

Thomas Borcherding, University of Washington: associate professor of economics, Virginia Polytechnic Institute and State University.

James M. Boughton: assistant professor of economics, Indiana University.

David A. Bowers, Case Western Reserve University: visiting associate professor of economics, Rice University.

Garry N. Bowles: visiting lecturer of accounting, University of North Carolina at Chapel Hill.

Marcel Boyer: assistant professor of economics, York University, July 1, 1971.

William K. Brandt: associate in business, Graduate School of Business, Columbia University, July 1, 1971.

Christine Branson, Council of Economic Advisers: instructor of economics, Simmons College.

Eugene F. Brigham, University of Wisconsin: graduate research professor of finance, University of Florida, Sept. 1971.

Eric Brown, University of Wisconsin: instructor of economics, Simmons College.

J. Michael Brown: assistant professor of economics and finance, University of Mississippi.

Joseph M. Burns, University of California, Los Angeles: associate professor of economics, Rice University, July 1971.

Winston Bush, Washington University: assistant professor of economics, Virginia Polytechnic Institute and State University.

Harry F. Campbell: assistant professor, department of economics, University of British Columbia.

Barry Carin: research assistant, Treasury Board, Planning Branch, Ottawa, Sept. 1971.

David Cass: professor of economics, Graduate School of Industrial Administration, Carnegie-Mellon University, Sept. 1970.

Richard J. Cebula: assistant professor of economics, Ohio University, Sept. 1971.

Charlotte A. Chamberlain: instructor of economics, Northeastern University.

Lowell C. Chesborough, University of Minnesota: assistant professor of real estate and urban land studies, University of Florida, winter 1972.

Barry M. Chiswick, Columbia University: associate professor of economics, Queens College of the City University of New York.

Victor Cholewicki: assistant professor, department of economics, Boston University.

Krishan Chopro: assistant professor of economics, State University of New York College of Oneonta.

William B. Clatanoff: assistant professor, department of economics, Mary Washington College, University of Virginia.

Robert W. Clower, Northwestern University: professor, department of economics, University of California, Los Angeles.

Benjamin J. Cohen: Woodrow Wilson School, Princeton: associate professor of international economic relations, Fletcher School of Law and Diplomacy, Tufts University, Sept. 1971.

Michael Conroy, University of Illinois at Urbana-Champaign: assistant professor, department of economics, University of Texas at Austin.

Jerry L. Cromwell: Instructor of economics, Northeastern University.

Dipankar Dasgupta: assistant professor of economics, York University, July 1, 1971.

Wilfred L. David: assistant professor, department of economics, Brooklyn College.

Benjamin G. Davis, University of Maryland: research economist, Resource Management Corporation, Feb. 1971.

Harold Demsetz, University of Chicago: visiting professor of economics, University of California, Los Angeles.

Warren Dent, University of Rochester: assistant professor, department of economics, University of Iowa.

Calvin L. DePass: instructor, department of economics, The American University.

Paul DeRosa: lecturer of economics, Douglass College, Rutgers University.

Arthur S. DeVany, University of California, Los Angeles: visiting associate professor, Texas A&M University.

Harry M. Douty: visiting professor, department of labor economics and income security, School of Industrial and Labor Relations, Cornell University.

Rolf H. Dumke: assistant professor, department of economics, University of British Columbia.

Richard Duvick, Economic Research Service, U.S.D.A.: assistant professor, agricultural economics department, Ohio State University, Aug. 1, 1971.

Richard Elnicki, Yale University: associate professor of management, University of Florida, Sept. 1971.

Richard D. Emmerson: assistant professor of economics, University of California, San Diego, 1971–72.

Noel J. J. Farley: associate professor, department of economics, Bryn Mawr College.

Allan M. Feldman, Johns Hopkins University: assistant professor, department of economics, Brown University, fall 1971.

Ann P. Fisher: assistant professor, department of economics, State University of New York College at Fredonia, August 29, 1971.

Frank R. Flatters: assistant professor of economics, Queen's University.

Eric Flamholtz: assistant professor, Graduate School of Business, Columbia University, July 1, 1971.

Russell Fogler, Brooklyn Polytechnic Institute: assistant professor of management, University of Florida, Sept. 1971.

Michele Fratianni: assistant professor of business economics, School of Business, Indiana University, Aug. 1971.

Alan Freiden, University of Chicago: assistant professor of economics, Virginia Polytechnic Institute and State University.

Richard T. Froyen: assistant professor of economics, University of North Carolina at Chapel Hill.

Earl I. Fuller, University of Massachusetts: professor and extension economist farm management, agricultural and applied economics; extension service, University of Minnesota.

A. Ronald Gallant: visiting assistant professor of economics and statistics, North Carolina State University.

Hugh B. Garnett, Jr., University of California, Berkeley: assistant professor of economics, Southern Methodist University, Sept. 1971.

Paul Gilbarg, University of Minnesota: professor, department of economics, Lewis and Clark College.

Curtis L. Gilroy: assistant professor, department of economics, Wells College.

Irvin N. Gleim, University of Illinois: assistant professor of accounting, University of Florida, Sept. 1971.

Andrew J. Gold: associate professor of economics and director, urban and environmental studies, Trinity College.

Lawrence Goldberg: research assistant, Human Resource Research Center, University of Southern California, Sept. 1971.

Claudia Goldin: assistant professor, department of economics, University of Wisconsin-Madison.

Philip E. Graves, Northwestern University: assistant professor of economics, Arizona State University.

James M. Griffin: assistant professor of economics, University of Houston.

Paul T. Gross, Jr., University of Rhode Island: associate professor of economics and business administration, University of Vermont.

Asher Halperin: visiting professor, Graduate School of Business, Columbia University, June 1, 1971.

Thomas Hamer: visiting assistant professor of economics, Bowling Green State University.

Masanori Hashimoto, Columbia University and National Bureau of Economic Research: assistant professor, department of economics, Wayne State University, Sept. 1971.

Dennis Henderson, Michigan State University: assistant professor, agricultural economic department, Ohio State University, Sept. 1, 1971.

J. Vernon Henderson: assistant professor of economics, Queen's University.

Edward A. Hewett, University of Michigan: assistant professor, University of Texas at Austin.

P. J. Hill: visiting assistant professor, department of economics, University of Iowa.

Fred Hitzhusen, Cornell University: assistant professor, agricultural economics department, Ohio State University, Aug. 1971.

Thomas Hogarty, Northern Illinois University: assistant professor of economics, Virginia Polytechnic Institute and State University.

Kenneth W. Hollman: assistant professor of economic research, University of Mississppi.

Charles R. Hulten: visiting assistant professor of political economy, Johns Hopkins University.

Robert B. Hume: instructor, department of economics, University of Rhode Island, July 1, 1971.

Michael D. Hurd: assistant professor, department of economics, Stanford University.

Peter Hutchinson: visiting assistant professor of economics, Bowling Green State University.

John Iton, McGill University: associate professor of economics, University of Pittsburgh, Sept. 1971.

Young S. Jang, State University of New York at Plattsburgh: assistant professor of economics and management, State University of New York at Albany.

William H. Jean: visiting professor of business administration, University of North Carolina at Chapel Hill, fall semester 1971.

Tim D. Kane: assistant professor of economics, Fort Hays Kansas State College.

Charles C. King: assistant professor of economics, Emory and Henry College, Sept. 1971.

Robert E. Klippel, Harvard University: assistant professor of marketing, University of Florida, Sept. 1971.

Jay E. Klompmaker: assistant professor of business administration, University of North Carolina at Chapel Hill.

Richard V. Knight: instructor, department of economics, Case Western Reserve University, Sept. 1971.

Robert T. Kudrle, Harvard University: assistant professor of economics, Texas A&M University.

Oldrich Kyn, University of California, Berkeley: associate professor, department of economics, Boston University.

Walter C. Labys, United Nations Conference on Trade and Development: assistant professor, department of economics, University of Rhode Island, July 1, 1971.

James A. Largay, Cornell University: assistant professor of accounting, Rice University, July 1971.

Judith R. Lave: assistant professor of economics, Graduate School of Industrial Administration, Carnegie-Mellon University, Sept. 1, 1971.

Arthur S. Lazare: assistant professor, department of economics, Brooklyn College.

Takyun J. Lee: assistant professor of economics, Lebanon Valley College, Sept. 1971.

Louis Lefeber: visiting professor of economics, York University, July 1, 1971.

Robert A. Lewis: assistant professor of economics, Ohio University, Sept. 1, 1971.

Lee A. Lillard III: instructor of economics, North Carolina State University.

David E. Lindsey: assistant professor of economics, Macalester College, fall 1971.

Charles D. Liner: assistant professor of economics, public law and government, University of North Carolina at Chapel Hill.

Irene Lurie: assistant professor, department of economics, Institute for Research on Poverty, University of Wisconsin-Madison.

Jerolyn R. Lyle: assistant professor, department of economics, The American University.

Gary L. McDowell: assistant professor of economics, Lewis and Clark College.

James B. McNamara: assistant professor, department of economics, Brooklyn College.

David L. McNicol, Harvard University: assistant professor, department of economics, University of Pennsylvania.

Paul M. Manchester: assistant professor, department of economics, Mary Washington College, University of Virginia.

Richard B. Mancke, University of Chicago: assistant professor, department of economics, University of Michigan.

Stanley H. Master, University of Wisconsin: visiting associate professor, department of economics, University of Notre Dame, Jan. 1972.

Michael Mazis, Pennsylvania State University: assistant professor of marketing, University of Florida, Mar. 1971.

David Meiselman, Macalester College: professor of economics, Virginia Polytechnic Institute and State University, Reston Campus.

Peter Mieszkowski: professor of economics, Queen's University.

Gerald M. Miller: instructor of economics, Miami University, Ohio, Sept. 1971.

Kendrick W. Miller, Jr., University of California, Berkeley: assistant professor of economics, University of California, Santa Cruz.

Richard K. Miller: instructor, department of economics, Washington and Jefferson College.

David Minars: assistant professor, department of economics, Brooklyn College.

Jora R. Minasian: visiting professor of economics, University of Southern California.

John A. Miranowski: instructor of economics, Northeastern University.

Herbert Mohring: professor of economics, York University, July 1, 1971.

Richard D. Morgenstern, American University: assistant professor of economics, Queens College of the City University of New York.

Samuel A. Morley, University of Wisconsin: visiting assistant professor of economics, Rice University.

Clarence C. Morrison: professor of economics. Indiana University.

M. H. Morse: professor, department of economics, Dalhousie University, July 1, 1971.

Laurence S. Moss, Columbia University: assistant professor of economics, University of Virginia, Sept. 1971.

Ernest A. Nagata, Purdue University: assistant professor of finance, University of Florida, Sept. 1971.

Carl W. Noller: assistant professor of economics, College of Business Administration, Valparaiso University, fall 1971.

Ken Norrie: assistant professor of economics, University of Alberta.

David Novack, Vassar College: lecturer of economics, State University of New York at Albany.

David J. Nye, University of Pennsylvania: assistant professor of insurance, University of Florida, Sept. 1971.

Gerald W. Olson: instructor, department of economics, University of Missouri at Kansas City, Sept. 1971.

Franklin L. Orth, Jr.: assistant professor of economics, University of Alaska.

Edward Osborne: assistant professor, department of economics, Marietta College.

Donald G. Paterson: assistant professor, department of economics, University of British Columbia.

Harold Payson III: instructor of international economic relations, Fletcher School of Law and Diplomacy, Tufts University, June 1971.

Carlos M. Pelaez: assistant professor of economics, Vanderbilt University.

Edmunds S. Phelps, University of Pennsylvania: professor of economics, Columbia University.

Edward S. Phillips, University of Missouri-Kansas City: assistant professor of economics, Marshall University.

David Pingry, Purdue University: assistant professor of economics, Virginia Polytechnic Institute and State University.

Charles Plott, Purdue University: professor of economics, California Institute of Technology.

Edward V. Prescott: assistant professor of economics, Graduate School of Industrial Administration, Carnegie-Mellon University, Sept. 1, 1971.

John J. Pringle: associate professor of business administration, University of North Carolina at Chapel Hill.

Douglas D. Purvis: assistant professor of economics, Queen's University.

Robert C. Radcliffe, University of Ohio: assistant professor of finance, University of Florida, Sept. 1971.

Akundi S. Rao, University of Rochester: assistant professor, department of economics, City College of the City University of New York.

Benjamin U. Ratchford: visiting professor of finance, University of Florida, winter quarter, 1971–72.

Michael Reich : assistant professor, department of economics, Boston University.

Joseph D. Reid, Jr., University of Chicago: assistant professor, department of economics, University of Pennsylvania.

Stephen Resnick, Yale University : associate professor, department of economics, City College of the City University of New York.

John Riley, Massachusetts Institute of Technology: instructor, department of economics, Boston College.

Hugh T. Rockoff: lecturer, department of economics, Rutgers-The State University, July 1971.

Joel B. Rosenberg: instructor of economics, State University of New York College at Geneseo.

Thomas Rossi: assistant professor of economics, State University of New York College at Oneonta.

David S. Rubin: assistant professor of business administration, University of North Carolina at Chapel Hill.

Brian S. Rungeling: assistant professor of economic research; codirector of Mid-South Center for Manpower Studies, University of Mississippi.

Yasuhiro Sakai: assistant professor of economics, University of Pittsburgh, Sept. 1971–73.

Gerald L. Salamon, Ohio State University: assistant professor of accounting, University of Florida, Sept. 1971.

Nicolas Sanchez, University of Southern California: assistant professor of economics, Texas A&M University.

Warren C. Sanderson: assistant professor, department of economics, Stanford University.

Todd M. Sandler, State University of New York: assistant professor of economics, Arizona State University.

Anthony Santomero: assistant professor of economics, City College of the City University of New York, Sept. 1971.

W. Phillip Saunders, Jr.: professor of economics, Indiana University.

Pawan K. Sawhney: instructor of economics, Northeastern University.

Mark E. Schaefer: assistant professor of economics, Georgia State University.

Richard M. Scheffler: assistant professor of economics and research associate, Health Services Center, University of North Carolina at Chapel Hill.

Peter J. Schmidt: assistant professor of economics, University of North Carolina at Chapel Hill.

Maria Schmundt: assistant professor, department of economics, University of Wisconsin-Madison.

S. Prakash Sethi: visiting assistant professor, Graduate School of Business, Columbia University, June 1, 1971.

James Shepherd, University of British Columbia: assistant professor of economics, Whitman College.

C. Richard Shumway: assistant professor of economics, North Carolina State University.

Margaret C. Simms, Stanford University: assistant professor of economics, University of California, Santa Cruz.

Frank A. Sloan, RAND Corporation and University of California, Los Angeles: assistant professor of economics, University of Florida, Sept. 16, 1971.

Dennis E. Smallwood, Cowles Foundation, Yale University: assistant professor of economics, University of California, San Diego.

Emmett D. Smith, Ohio State University: assistant professor of accounting, University of Florida, Mar. 1971.

Halbert C. Smith, Ohio State University: visiting professor of real estate and urban land studies, University of Florida.

Kenneth R. Smith, University of Wisconsin: visiting associate professor of economics, University of California.

Lewis H. Smith: assistant professor of economic research, University of Mississippi.

John J. Solady: assistant professor of economics, Pennsylvania State University, Sept. 1971.

Clive Southey, University of British Columbia: assistant professor, department of economics, University of Guelph, July 1, 1971.

Miron Stano, Cornell University: assistant professor, department of economics, Wayne State University, Sept. 1971.

Paula E. Stephan: assistant professor of economics, Georgia State University.

Dana N. Stevens: assistant professor of economics, Georgia State University.

Douglas B. Stewart, University of Oregon: visiting lecturer, department of economics, San Diego State College.

Mahlon R. Straszheim, Harvard University: associate professor of economics, University of Maryland.

Raymond J. Struyk, U.S. Air Force: visiting assistant professor of economics, Rice University.

David H. Stuart: instructor of economics, North Carolina State University.

William C. Stubblebine, Claremont Men's College: visiting associate professor of economics, Southern Methodist University, Sept. 1971.

Andrew M. Sum: instructor of economics, Northeastern University.

Dalip S. Swamy, Standard Oil Company, New Jersey: visiting associate professor of economics, Southern Methodist University, Sept. 1971.

William K. Tabb, University of Connecticut: associate professor of economics, Queens College, City University of New York.

Leonard J. Tashman, Federal Reserve Board: assistant professor of economics and business administration, University of Vermont.

Russell Taussig: visiting professor, Graduate School of Business, Columbia University, June 1, 1971.

Dean G. Taylor, University of Chicago: assistant professor of finance, University of Florida, Sept. 1971.

Ronald J. Teichman, Northwestern University: assistant professor of accounting, University of Florida, Sept. 1971.

Thom B. Thurston: assistant professor of economics, Queens College, City University of New York.

Frank Tinari: assistant professor of economics, Seton Hall University, Sept. 1971.

Thomas R. Tirney, Temple University: assistant professor of economics and business administration, University of Vermont.

Bonne A. Turchi: assistant professor of economics and research associate, Population Program, University of North Carolina at Chapel Hill.

Aldon A. Turner: instructor of economics, North Carolina State University.

Adolph L. Vandendorpe: associate professor of economics, Macalester College, fall 1971.

John M. Volpe: assistant professor, department of economics, Connecticut College.

George M. von Furstenberg: associate professor of economics, Indiana University.

Terence J. Wales, University of Pennsylvania: associate professor, department of economics, University of British Columbia.

Gary M. Walton: professor of economics, Indiana University.

Thomas Walton: assistant professor of economics and finance, Graduate School of Business, New York University.

Stanley Warner: professor of administrative studies and economics, York University, July 1, 1971.

John Wassom, University of Florida: associate professor of economics, Western Kentucky University.

Norman L. Weed: associate professor, department of economics, Wichita State University, Sept. 1971.

Merle Y. Weiss, National Bureau of Economic Research, Inc.: lecturer, department of economics, University of Pennsylvania.

Francine Weisskoff: instructor of economics, Trinity College.

William S. Westbrook, Ohio University: associate professor of economics, Marshall University.

Kenneth Wieand: assistant professor of economics, Marquette University.

Albert R. Wildt, Purdue University: assistant professor of marketing, University of Florida, Sept. 1971.

Alan Winger, University of Kentucky: professor of economics and urban studies, University of Florida, Sept. 16, 1971.

Benjamin W. Wolingson: assistant professor, School of Labor and Industrial Relations, Michigan State University, Sept. 1971.

Gopal J. Yadav: assistant professor, department of economics, University of British Columbia.

Richard Yates: assistant professor of economics, State University of New York College at Onenota.

Leaves for Special Appointments

Ruben Almonacid, Vanderbilt University: University of Sao Paulo, Brazil.

Edward I. Altman, New York University: visiting professor of finance, Institute Superieur des Affaires, Ecoledes Etades Commerciales, Jouy-en-Josas, France.

Nathan A. Baily, American University: commissioner, Postal Rate Commission, Washington.

H. Walter Baumgartner, State University of New York College at Oneonta: University of Wurtzburg, Germany.

Jere R. Behrman, University of Pennsylvania: Ford Foundation Faculty Research Fellowship.

Stanley M. Besen, Rice University: Brookings Economic Policy Fellow, Office of Telecommunications Policy.

Kenneth M. Brown, University of Notre Dame: visiting lecturer, University of Nairobi, 1971–73.

Edwin Burmeister, University of Pennsylvania: Duke University.

William K. Chung, Denison University: guest scholar, Center for Advanced Study, Brookings Institution.

Carolyn Clark, University of British Columbia: visiting assistant professor, department of economics, Queen's University.

Kenneth W. Clarkson, University of Virginia: Public Policy Fellow, Office of Management and Budget, Brookings Institution.

John J. Dall, Seton Hall University: State of New Jersey.

J. D. DeForest, Economic Development Administration: Office of Environmental Affairs, U.S. Department of Commerce, Washington.

William P. Dillingham, Florida State University: visiting professor, Florida State University Center, Panama Canal Zone, spring and summer 1971.

Noel Edelson, University of Pennsylvania: Tel Aviv University, Israel.

Robert Eisner, Northwestern University: visiting professor of economic policy, department of economics, State University of New York at Binghamton, fall 1971.

Ronald E. Findlay, Columbia University: Ford Faculty Fellow.

William E. Gibson, University of California, Los Angeles: senior staff economist, Council of Economic Advisers.

Michael Greenwood, Kansas State University: Brookings Economic Policy Fellowship, Economic Development Administration, U.S. Department of Commerce.

Thomas H. Havrilesky, Rice University: visiting associate professor, department of economics and commerce, Simon Fraser University, summer 1971.

Janos Horvath, Butler University: senior Fellow, Research Institute on Communist Affairs, Columbia University.

James P. Houck, University of Minnesota: Kasetsart University, Bangkok.

Edwin J. Kane, Boston College: visiting professor, department of economics, Simon Fraser University, summer 1971.

Kwan S. Kim, University of Notre Dame: visiting lecturer, University of Nairobi, 1971–73.

Benjamin Klein, University of California, Los Angeles: research Fellow, National Bureau of Economic Research.

Joseph S. La Cascia, Marshall University: visiting professor, Senior Fulbright-Hays Fellow, University of Vera Cruz, Mexico.

David T. Lapkin, University of North Carolina at Chapel Hill: visiting professor, Haifa University, Israel.

Robert Larner, Brandeis University: industrial organization economist, Bureau of Economics, Federal Trade Commission, Washington.

Robert Lindsay, New York University: head of staff, New York State Council of Economic Advisers.

Cliff Lloyd, University of Iowa: visiting professor, department of economics and commerce, Simon Fraser University, summer 1971.

Bevars D. Mabry, Bowling Green State University: visiting professor of economics, Rockefeller Foundation, Thammasat University, Bangkok.

M. C. Madhavan, San Diego State College: visiting professor of economics, National Institute of Bank Management, Bombay, India.

H. Michael Mann, Boston College: director, Bureau of Economics, Federal Trade Commission.

Edwin Mansfield, University of Pennsylvania: Center for Advanced Study in the Behavioral Sciences, Stanford University.

G. H. Mattersdorff, Lewis and Clark College: National Science Foundation, London School of Economics and Political Science.

Gordon R. Munro, University of British Columbia: visiting professor, School of Comparative Social Sciences, University of Penang, Gelugor, West Malaysia.

J. Carter Murphy, Southern Methodist University: senior staff economist, Council of Economic Advisers.

Woo H. Nam, San Diego State College: senior economist, Korean Development Institute.

Almarin Phillips, University of Pennsylvania: President's Commission on Financial Structure and Regulation and University of Warwick, England.

Robert H. Rasche, University of Pennsylvania: Federal Reserve Bank of St. Louis.

Il SaKong, New York University: visiting lecturer, University of Sheffield, England.

Calvin D. Siebert, University of Iowa: visiting professor, University of Philippines.

Robert L. West, Fletcher School of Law and Diplomacy, Tufts University: visiting professor of international economics, Graduate Institute of International Sudies, Geneva, Switzerland, spring semester 1971.

Edward A. Wiegner, Marquette University: Secretary of Revenue, State of Wisconsin.

Jay W. Wiley, Purdue University: The London School of Economics and Political Science, Apr.-July 1971.

John H. Young, University of British Columbia: chairman, Prices and Incomes Commission, Ottawa.

Resignations

Arthur A. Bayer, University of Vermont.

Diran Bodenhorn, Ohio State University; Cleveland State University, Sept. 1971.

Robert C. Bushnell, Ohio State University: Wayne State University, Sept. 1971.

Rondo Cameron, University of Wisconsin-Madison: Emory University.

V. Karuppan Chetty, Columbia University: Center for Operations Research and Econometrics, Université Catholique de Louvian.

Karen Davis, Rice University: Brookings Institution, July 1, 1971.

Alfred S. Eichner, Columbia University, June 1971.

Kenneth Frederick, California Institute of Technology: Resources for the Future.

Paul K. Gatons, Georgia State University: Louisiana Tech University, Sept. 1971.

Charles E. Gearing, University of North Carolina at Chapel Hill.

Berch Haroian, Seton Hall University: Dominican College.

Claude Hillinger, Case Western University: Georgia Insitute of Technology, June 1971.

Charles Holt, University of Wisconsin-Madison: Urban Institute.

Thomas A. Hoopengardner, Marshall University, May 31, 1971.

Donald W. Katzner, University of Pennsylvania: University of Waterloo, June 1971.

Alan P. Kirman, Johns Hopkins University: Center for Operations Research and Econometrics, Université Catholique de Louvain.

Hugh W. Knox, University of North Carolina at Chapel Hill.

Ben E. Laden, Ohio State University: Federal Reserve Board, Washington, Sept. 1971.

Charles R. Laudor, Connecticut College.

James W. McKie, Vanderbilt University: University of Texas, Austin.

Robert Maddox, University of Florida, June 1971.

Jeffrey H. Mandell, University of Florida, June 1971.

Emil M. Meurer, Jr., University of Nebraska: Mankato State College, Aug. 1971.

Brian Motley, University of Wisconsin-Madison.

Ranganathan Narayanan, Carnegie-Mellon University: St. Mary's University, Halifax.

David H. Nissen, Rice University: Resource Management Corporation, Bethesda, Sept. 1971.

Roger Noll, California Institute of Technology: Brookings Institution.

Thomas O'Brien, Brandeis University: University of Minnesota, June 1971.

Larry Orr, University of Wisconsin-Madison.

Kenneth M. Parzych, State University of New York at Fredonia: Nichols College, June 30, 1971.

Robert Pickhardt, University of Florida, Aug. 1971.

Clayne Pope, University of Michigan: Brigham Young University, Jan. 1971.

T. Ramamohan Rao, Kansas State University: Indian Technical Institute at Khanpur.

Thomas J. Sargent, University of Pennsylvania: University of Minnesota, June 1971.

Joseph F. Shaw, University of Southern California.

Jeannine Swift, State University of New York College at Geneseo.

Benjamin J. Taylor, Arizona State University: University of Oklahoma, July 15, 1971.

Emanuel Tobier, Queens College of the City University of New York: New York University, June 1971.

Stephen J. Turnovsky, University of Pennsylvania: University of Toronto, June 1971.

Gordon Tush, University of Nebraska at Omaha.

Eldon E. Weeks, Washington State University: Farm Income Branch, Economic and Statistical Analysis Division, ERS, Washington.

Norman E. Weir, University of Mississippi.

Neil Weiss, Graduate School of Business, Columbia University, June 30, 1971.

Phillip M. Weitzman, New York University, June 1971.

Janet M. Whisler, Mary Washington College, University of Virginia.

Edward E. Williams, Rutgers-The State University: McGill University, June 30, 1971.

Claus Wittich, University of Southern California.

Miscellaneous

Rawle Farley: elected chairman, State University of New York Faculty Association for African Studies, 1971–72.

The American Economic Review

[illegible]

THE AMERICAN ECONOMIC ASSOCIATION

Founded in 1885

• Published at George Banta Co., Inc., Menasha, Wisconsin.

• The American Economic Review, including four quarterly numbers, the *Proceedings* of the annual meetings, and *Directory* and Supplements, is published by the American Economic Association and is sent to all members five times a year, in March, May, June, September and December.

• Membership dues of the Association are $20.00 a year, which includes a year's subscription to both the *American Economic Review* and the *Journal of Economic Literature.* Subscriptions by nonmembers are $20.00 a year, and only subscriptions to both publications will be accepted. Single copies of the *Review* and *Journal* are $4.00 each. Each order for copies of either publication must also include a $.50 per order service charge. Orders should be sent to the Secretary's office, Nashville, Tennessee.

• Correspondence relating to the *Paper and Proceedings*, the *Directory,* advertising, permission to quote, business matters, subscriptions, membership and changes of address may be sent to the secretary, Rendigs Fels, 1313 21st Avenue, South, Nashville, Tennessee 37212. To be effective, notice of change of address must reach the secretary by the 1st of the month previous to the month of publication.

• Second-class postage paid at Nashville, Tennessee and at additional mailing offices. Printed in U.S.A.

Officers

President

JAMES TOBIN
Yale University

President-Elect

JOHN KENNETH GALBRAITH
Harvard University

Vice-Presidents

JAMES M. BUCHANAN
Virginia Polytechnic Institute
FRANCO MODIGLIANI
Massachusetts Institute of Technology

Secretary-Treasurer and Editor of Proceedings

RENDIGS FELS
Vanderbilt University

Managing Editor of The American Economic Review

GEORGE H. BORTS
Brown University

Managing Editor of The Journal of Economic Literature

MARK PERLMAN
University of Pittsburgh

Executive Committee

Elected Members of the Executive Committee

MARY JEAN BOWMAN
University of Chicago
CHARLES L. SCHULTZE
Brookings Institution
ROBERT DORFMAN
Harvard University
ARNOLD C. HARBERGER
University of Chicago
ROBERT EISNER
Northwestern University
JOHN R. MEYER
Yale University

Ex Officio Members

WILLIAM FELLNER
Yale University
WASSILY LEONTIEF
Harvard University

Money and Barter in General Equilibrium with Transactions Costs

By Jürg Niehans*

In this paper, the neoclassical approach to monetary theory[1] is generalized by the explicit introduction of transactions costs. The neoclassical theory of general monetary equilibrium is based on the twin assumptions that 1) the economy is monetized, and 2) the utility of money can be treated *as if* it were of the same kind as the utility of consumer goods or factors of production. The theory does not permit us to explain why an economy uses money rather than barter, nor how it chooses its money. It also fails to provide an explicit analysis of the services of money in saving resources used up in the exchange process. It is often argued, therefore, that money is essentially a dynamic phenomenon, related to speculation and uncertainty, which can not be analyzed satisfactorily in terms of equilibrium theory.[2] However, there is always the qualification that this argument is true only in the absence of transactions costs. This paper is written from the complementary point of view that a theory of money, valid for economic equilibrium, can indeed be built on the concept of transactions costs. It will be shown that this theory is able to explain the emergence of different exchange arrangements, ranging from barter to full monetization, as the result of differences in transactions costs. It also gives an explicit analysis of the (indirect) utility of money arising from the exchange process. The discussion will be restricted to a pure exchange economy. The exchange and storage of goods will be the only production activities, and there is no credit, interest, or time preference.[3]

I. General Features of the Model

Let time be divided into years and days. Each year is assumed to be the same as any other; in this respect the model is stationary, excluding capital accumulation and growth. Within each year things are subject to change from day to day; to this extent the model is dynamic. This dynamic element provides a rationale for the holding of stocks, including cash balances, making it possible to formalize the notion that money is held to "bridge the gap" between receipts and payments.

Individuals are assumed to be endowed with certain quantities (sometimes zero) of each good. These quantities can be consumed, exchanged, or stored. Exchange takes place in different "markets."[4] We cannot start by assuming, however, that in each market a certain commodity is traded against money, because this is what eventually we want to explain, namely full monetization. We rather have to assume that there is a separate trading post for

---

* Department of political economy, Johns Hopkins University. The financial support by National Science Foundation Grant GS-2217 is gratefully acknowledged.

[1] See the recent restatement by Paul A. Samuelson.

[2] See John R. Hicks, P. N. Rosenstein-Rodan.

[3] The present model is a further development of the approach I used in my 1969 and 1970 articles. While consumption was there assumed to be given independently of monetary arrangements, the two are now jointly determined.

[4] The use of the market concept is not really necessary. The analysis could well be conducted in terms of exchange between individuals (see, for example, my 1969 and 1970 articles). Of course, an individual exchange theory is richer in economic content than a market theory. The latter approach was chosen here because it simplifies the notation and corresponds more closely to the neoclassical approach.

each pair of goods. At one post, apples are traded against nuts, at another apples are exchanged against pears, and so on.[5] For q goods, there are thus $(q/2)\ (q-1)$ markets. Under given conditions, some of these markets will usually turn out to be inactive. In a fully monetized economy, all $q-1$ nonmonetary commodities are exchanged against money only; the remaining $(1/2)\ (q-1)\ (q-2)$ trading posts are deserted.

It is an essential feature of the model that exchange causes costs. Transactions costs are defined as those costs which arise not from the production of goods (including transportation, packing, etc.) but from their transfer from one owner to another.[6] It will be assumed for simplicity that all transactions costs are expressed in terms of "work," measuring the time and trouble in making transactions. In principle, it would be possible to specify for each transaction a production function relating the transfer of a certain good to the required inputs of many factors. For the present purpose a one-factor production function is sufficient. In part, transactions costs vary with the quantity transferred: The transfer of two houses may involve almost twice the costs of the transfer of one. Another part may be fixed: The resource costs of a stock transaction is little affected by the number of shares. I showed in an earlier paper[7] that fixed transactions costs give rise to an integer programming problem whose shadow prices have no straightforward interpretation in terms of competitive prices. We cannot expect, therefore, that the market mechanism provides for an optimal exchange arrangement. While the existence of fixed transactions costs may thus have far-reaching consequences for the welfare implications of general equilibrium, there is nothing that could be said at the present time beyond what was already indicated in the above mentioned paper. In the present context fixed transactions costs will therefore be disregarded. The following assumptions are made with respect to variable transactions costs:

1) They are counted separately for the two goods exchanged in a given transaction with no interdependence between the two sides;
2) they depend both on the transactor and the good;
3) they are the same for goods bought and sold;
4) they do not depend on the aggregate volume of transactions in a given market, thus excluding a potentially important externality.

There are also costs for the holding of stocks, reflecting wastage, maintenance, protection, loss, etc.[8] These storage costs introduce another element of production into the model, for which we have to specify a technology. Again it would be possible to specify production functions relating storage to different resource inputs. For the present purposes we shall assume that storage costs consist simply in a shrinkage of stocks by a certain percentage per period. This rate may be different for different individuals and goods, but it is assumed to be constant over time. It can be interpreted as an own-rate of interest to be paid for the holding of a certain stock.[9] This conceptualization of holding costs means that capital is visualized as working capital in the form of inventories, while problems of time preference, interest, credit, and capital goods are not touched upon. Correspondingly, any demand for money results from transactions motives.

---

[5] In principle, there could also be trading posts for triples, quadruples, . . . n-tuples of goods. It is assumed that transactions costs on such higher-order markets are prohibitive—as in reality they usually are.

[6] The term is a catchall for different items. Costs may arise because the parties have to communicate, the goods have to be inspected, measured and marked, a title search may be necessary, contracts have to be drawn up, title of ownership has to be transferred in certain prescribed forms, the transfer may have to be recorded etc. There are good reasons for classifying many of these items as search or information costs (see Karl Brunner and Allan H. Meltzer). In the present context it is both impossible and unnecessary to introduce these uncertainty elements in an explicit way.

[7] See Niehans (1969, pp. 722 ff).

[8] In a more complete analysis, these costs would also include subjective time preference, while a part of them would be offset by positive yields, reflecting in some cases (like trees) physical growth, in others (like bonds) interest payments. The sum of these components will have to be a negative yield or a positive cost; otherwise there is no upper limit to steady-state stocks.

With given endowments, market prices, transactions costs and storage costs, individuals are supposed to schedule their exchange transactions and stocks in such a way that the utility of the resulting consumption pattern is maximized. General equilibrium will be reached when prices are such that demand and supply are equal (though possibly zero) at every trading post. The question is under what conditions this equilibrium will be one of barter and under what circumstances a good will emerge in the role of money.

II. Individual Optimization

The model outlined in the preceding section will now be specified in detail. We begin with the optimization problem of an individual faced with given market conditions. Suppose there are $1 \ldots h,\ k \ldots q$ goods and $1 \ldots \mathrm{t} \ldots \mathrm{T}$ days, where transactions costs are assumed to be paid out of the qth good. The individual is endowed with resources ${}^{1}\bar{x}^{1} \ldots {}^{\mathrm{t}}\bar{x}^{h} \ldots {}^{\mathrm{T}}\bar{x}^{q}$. By trading in the market and carrying stocks from one day to another the individual maximizes the differentiable utility function of consumption x,

$$U = U({}^{1}x^{1}, \ldots, {}^{\mathrm{t}}x^{h}, \ldots, {}^{\mathrm{T}}x^{q}) \tag{1}$$

which is assumed to have convex isoquants and positive marginal utility for at least one commodity. The quantity of good h sold against k on day t will be denoted ${}^{\mathrm{t}}y^{hk}$, while ${}^{\mathrm{t}}z^{kh}$ is the quantity of k bought against h. Denote ${}^{\mathrm{t}}s^{h}$ to be the stock of good h held on day t and γ^{h} the daily holding costs per unit of such stocks. For each good except the q^{th}, the resource constraint states that consumption minus net purchases plus addition to stocks plus storage costs equals the endowment for each day of the year:

$$\begin{aligned} {}^{\mathrm{t}}x^{h} - \sum_{k}^{q} ({}^{\mathrm{t}}z^{hk} - {}^{\mathrm{t}}y^{hk}) + ({}^{\mathrm{t}}s^{h} - {}^{\mathrm{t}-1}s^{h}) \\ + \gamma^{h}\, {}^{\mathrm{t}}s^{h} = {}^{\mathrm{t}}\bar{x}^{h} \qquad (h = 1 \ldots q-1;\ k \neq h) \end{aligned}$$

Purchases and sales at a given trading post are related through the exchange ratio or price

$${}^{\mathrm{t}}p^{hk} = \frac{{}^{\mathrm{t}}z^{kh}}{{}^{\mathrm{t}}y^{hk}} = \frac{{}^{\mathrm{t}}y^{kh}}{{}^{\mathrm{t}}z^{hk}} = 1/{}^{\mathrm{t}}p^{kh}$$

Substituting for ${}^{\mathrm{t}}z^{hk}$, the $\mathrm{T}(q-1)$ resource constraints can be written

$$\begin{aligned} {}^{\mathrm{t}}x^{h} = \sum_{k}^{q} ({}^{\mathrm{t}}p^{kh}\, {}^{\mathrm{t}}y^{kh} - {}^{\mathrm{t}}y^{hk}) \\ + ({}^{\mathrm{t}}s^{h} - {}^{\mathrm{t}-1}s^{h}) + \gamma^{h}\, {}^{\mathrm{t}}s^{h} = {}^{\mathrm{t}}\bar{x}^{h} \end{aligned} \tag{2}$$

The T resource constraints for q include additional terms reflecting transactions costs. If c^{h} is the cost, measured in units of q, of transferring one unit of good h, these constraints can be written

$$\begin{aligned} {}^{\mathrm{t}}x^{q} - \sum_{k}^{q-1} ({}^{\mathrm{t}}p^{kq}\, {}^{\mathrm{t}}y^{kq} - {}^{\mathrm{t}}y^{qk}) \\ + ({}^{\mathrm{t}}s^{q} - {}^{\mathrm{t}-1}s^{q}) + \gamma^{q}\, {}^{\mathrm{t}}s^{q} \\ + \sum_{h}^{q} \sum_{k \neq h}^{q} (c^{h} + c^{k}\, {}^{\mathrm{t}}p^{hk})\, {}^{\mathrm{t}}y^{hk} = {}^{\mathrm{t}}\bar{x}^{q} \end{aligned} \tag{3}$$

Since, for the year as a whole, this is intended to be a stationary model, terminal stocks are required to be equal to initial stocks:

$${}^{\mathrm{t}}s^{h} = {}^{0}s^{h} \qquad (h = 1 \ldots q)$$

---

[9] For services (including work) storage costs must be assumed to be infinite or in any case prohibitive.

$$L = U({}^1x^1, \ldots, {}^{\mathrm{T}}x^q) - \sum_{t}^{\mathrm{T}} \sum_{h}^{q} {}^t\lambda^h \left[{}^tx^h - \sum_{k}^{q} ({}^tp^{kh}\, {}^ty^{kh} - {}^ty^{hk}) + {}^ts^h - {}^{t-1}s^h + \gamma^h\, {}^ts^h - {}^t\bar{x}^h \right]$$

$$- \sum_{t}^{\mathrm{T}} {}^t\lambda^q \left[\sum_{h}^{q} \sum_{k}^{q} (c^h + c^k\, {}^tp^{hk})\, {}^ty^{hk} \right] \qquad (h, k = 1 \ldots q; k \neq h)$$

In addition, all flows and stocks must be nonnegative:

$${}^tx^h \geqq 0, \quad {}^ty^{hk} \geqq 0, \quad {}^ts^h \geqq 0 \quad (h = 1 \ldots q)$$

It should be emphasized that initial stocks, like other stocks, are not given to the individual, but determined endogenously by the optimizing calculus. The model thus determines those initial stocks which, once they are attained at the beginning of a given year, give the individual no reason to plan for a change in initial stocks next year. The analysis of the approach to equilibrium stocks is outside the scope of this paper.

What we have described is a non-linear programming problem with a quasi-concave objective function and linear constraints. Its solution determines consumption, exchange flows and stocks, as functions of endowments, market prices, transaction costs, and storage costs. In symbols, using vector notation,

$$x = x(\bar{x}, p, c, \gamma)$$

$$y = y(\bar{x}, p, c, \gamma)$$

$$z = z(\bar{x}, p, c, \gamma)$$

$$s = s(\bar{x}, p, c, \gamma)$$

The problem is thus a straightforward generalization of the neoclassical problem of utility maximization under a budget constraint for a world with transactions and storage costs. It should be noted, however, that the resulting functions do not have a homogeneity property corresponding to that of neoclassical theory. Also, contrary to the neoclassical model, an individual may well be both a buyer and seller of the same good, though at differen trading posts, that is, against differen other goods. On the other hand, man trading posts may not be frequented b him at all.

For a problem of this type Kennet Arrow and Alain Enthoven have show that the Kuhn-Tucker conditions ar necessary and sufficient for a maximum These conditions are obtained from th Lagrange function shown above. The multiplier ${}^t\lambda^h$ can be interpreted as the marginal value or opportunity cost of resourc h on day t.[10] The Kuhn-Tucker condition relating to consumption are

$$(4) \qquad \left(\frac{\partial U}{\partial {}^tx^h} - {}^t\lambda^h \right) {}^tx^h = 0$$

$$\frac{\partial U}{\partial {}^tx^h} - {}^t\lambda^h \leqq 0 \qquad {}^tx^h \geqq 0$$

If a good is consumed at all, its margina utility in consumption is equalized to it marginal value or opportunity cost. Th inequality sign implies that a good, in vie of its services in exchange or as a store o wealth, may have a positive margina value even if it has no direct utility.

There is another set of conditions fo market supply[11]

---

[10] This statement is based on M. L. Balinski an William Baumol who also give the required qualific tions.

[11] The implied conditions for market demand can b obtained from the supply conditions together with th definition of prices. With respect to the first term of th supply conditions note that in the Lagrange functio ${}^tp^{hk}\, {}^ty^{hk}$ appears with multiplier ${}^t\lambda^k$.

$$(5)\quad \begin{aligned} \left[{}^t\lambda^k\,{}^tp^{hk} - {}^t\lambda^h - {}^t\lambda^q(c^h + c^k\,{}^tp^{hk})\right]{}^ty^{hk} &= 0 \\ {}^t\lambda^k\,{}^tp^{hk} - {}^t\lambda^h - {}^t\lambda^h(c^h + c^k\,{}^tp^{hk}) &\leqq 0 \\ y^{hk} &\geqq 0 \end{aligned}$$

The first inequality can be rearranged to read

$$(6)\qquad {}^tp^{hk} \leqq \frac{{}^t\lambda^h + {}^t\lambda^q c^h}{{}^t\lambda^k - {}^t\lambda^q c^k}$$

This is the counterpart to the familiar condition that relative prices are equal to the marginal rate of substitution.

For stocks, finally, we obtain

$$(7)\quad \begin{aligned} &\left[{}^{t+1}\lambda^h - {}^t\lambda^h(1+\gamma^h)\right]{}^ts^h = 0 \\ &{}^{t+1}\lambda^h - {}^t\lambda^h(1+\gamma^h) \leqq 0; \quad {}^ts^h \geqq 0 \end{aligned}$$

The first inequality can be written

$$(8)\qquad {}^{t+1}\lambda^h - {}^t\lambda^h \leqq \gamma^h\,{}^t\lambda^h$$

The expression on the left is the marginal value of good h held in stock for one day; in contrast to the marginal flow value ${}^t\lambda^h$ we may call it the marginal stock value.[12] The condition means that in equilibrium, if a stock is held at all, the marginal gain from holding it for a day is just equal to the marginal loss because of storage costs. If the gain is smaller, no stocks are held. Holding costs act as an effective brake on stock accumulation. If holding costs are zero, stocks are held at satiation levels, the marginal stock values go to zero, and marginal flow values ${}^t\lambda^h$ are equalized over time. For negative holding costs or positive yields there is no equilibrium; further accumulation always pays.[13]

III. Market Equilibrium

We now assume that we know the consumption, supply, demand, and stock functions for each of n individuals. The prices on the $(q/2)\ (q-1)$ markets for each of T days, on the other hand, become endogenous variables. These prices have to satisfy the requirement that aggregate demand equal aggregate supply for each good. Denoting individuals by subscripts, we thus obtain the set of $Tq(q-1)$ conditions

$$\sum_i^n {}^tz_i^{hk} = \sum_i^n {}^ty_i^{hk}$$

$$(t = 1 \ldots T;\ h, k = 1 \ldots q;\ h \neq k)$$

However, by the definition of prices, if this condition is satisfied for one of the goods traded at a given trading post, it is automatically satisfied for the other. Therefore, only half of the above equilibrium conditions are independent.[14] The number of independent equilibrium conditions thus matches the number of prices.

Once equilibrium prices are known, consumption, market demand and supply, and stocks for all individuals are determined simultaneously. The result is a general equilibrium of barter. It should be noted that this system has, in general, none of the familiar homogeneity properties found in neoclassical models. In particular, the system cannot be dichotomized into a "real" and a "monetary" part. It will be shown below how the neoclassical properties can be made to result from special assumptions.

It is important to realize that for some markets the equilibrium trade volume may be zero. For one day, two traders and two goods (and thus one market), this aspect can be illustrated graphically. The optimization problem of trader 1 is represented in Figure 1. Note that E is the no-trade point, defined by the given endowments. Trade can take place along the straight line through E whose slope measures the

---

[12] Alternatively, (8) could be rearranged to say that the marginal stock value, capitalized at the storage cost rate, must not exceed the marginal flow value.

[13] Mathematically this shows up in the fact that for $\gamma^h<0$, (8) cannot be satisfied for all time periods simultaneously without violating the stationarity condition.

[14] This is the counterpart to the neoclassical proposition that equilibrium in $n-1$ markets implies equilibrium in the nth market.

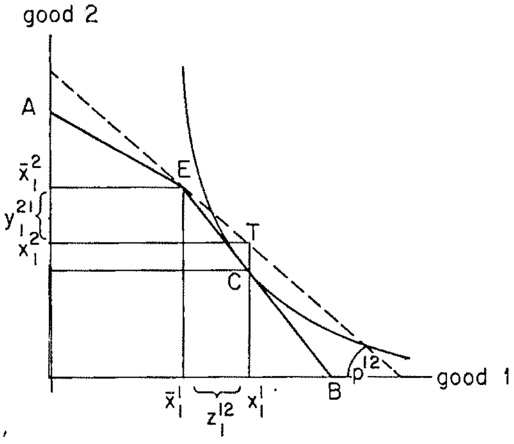


FIGURE 1

market price of good 1 in terms of good 2. However, in view of transactions costs the commodity bundle resulting from trade is not fully available for consumption. The consumption frontier is *AEB*. If transactions costs are paid out of good 2, they are measured by the vertical distance between the consumption frontier and the trade line. The trader will maximize utility by choosing consumption pattern *C*. This results in market demand and supply represented by the difference between *T* and *E*.

With a change in market price, both the trade line and the consumption frontier will swivel around *E*. If the initial consumption equilibrium was somewhere along the flat segments *AE* and *EB*, this will lead to a change in market demand and/or supply. However, if the equilibrium was initially at the no-trade point *E*, it is clear that it may stay there for some range of prices. A similar graph could be drawn for the second trader.

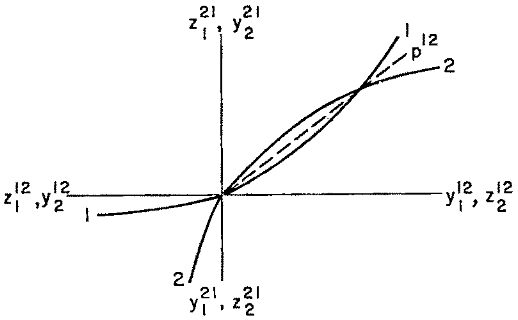


FIGURE 2

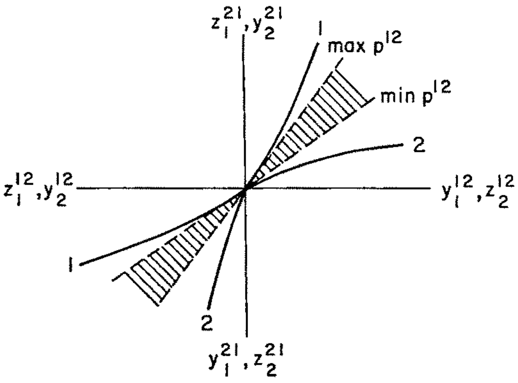


FIGURE 3

The results for market demand and supply are summarized by the offer curves for traders 1 and 2 in Figure 2, where the origin represents the no-trade point. The kink in each consumption frontier, due to transactions costs, is reflected in a kink of the respective offer curve at the origin. The intersection of these curves then determines the direction of trade, if any, the market price, and the quantities exchanged. It may happen, however, that market equilibrium is at the no-trade point, as illustrated in Figure 3. In this case, we can generally determine a range of prices only, any price within this (shaded) range making the desired exchange zero for both traders.[15]

This two-person one-market example shows that the no-trade solution is more likely, *ceteris paribus*, the higher the transactions cost rate. In a multimarket system, various sets of markets can be made inactive by choosing appropriate configurations of transactions cost rates. In one case, all markets may be active, implying that every good is traded against every other good. In another case, all markets are deserted except those where one particular good is traded against other goods; this would be the case of a fully monetized economy. From the point of view of explaining the emergence of different exchange arrangements, it is thus one of the principal functions of the model to determine which trading posts are active under a given set of conditions.

---

[15] As the figure is drawn, a p^{12} below min p^{12} or above max p^{12} would make the price line intersect the offer curve of trader 1 either in the third or first quadrant, respectively (but not necessarily the offer curve of 2).

IV. Barter and Commodity Money

Different exchange arrangements, some monetary, others nonmonetary, can be made to emerge by varying the pattern of transactions costs. In the present section, the timeless aspects of this proposition will be considered in the context of a one-period model. It is derived from the more general model presented above by simply omitting time superscripts and stocks.[16]

Suppose, with given tastes and endowments, we start out by assigning arbitrary positive values to transaction cost rates in a random fashion.The result is likely to be a complex network of exchange in which some trading posts will be active, others idle, including cases of both direct barter and indirect exchange. In the latter cases, goods are bought for resale, thus serving as media of exchange. The same individual may use several such media, each in a different range of applications, and different individuals will often use different sets of exchange media. There is no general reason to expect that any one good will universally or even widely be used as a medium of exchange and thus deserve the name of money. From this general case we shall now try to derive special (and in some sense limiting) cases by varying transactions costs in some particular way.

We first assume that wherever there is an indirect exchange, the transactions cost rate on the respective medium of exchange is progressively increased. It is clear that there is always a rate high enough to make it preferable to eliminate that medium of exchange, replacing indirect by direct barter.[17] In the course of this process we will sooner or later arrive at an exchange system with direct barter only. It is regarded as commonplace that such a system is wasteful. This is not generally true, though, because with the given exchange technology, embodied in a particular set of transactions cost rates, it may well be the most efficient trading pattern available.

In a second experiment we assume that transactions cost rates are increased wherever there is any exchange, be it direct or indirect. In general, this will tend to lower the volume of market exchange. This is so, in part, because the rise in transactions costs, similar to a tax on exchange, will make it relatively more profitable to consume one's own endowment. A possible result is complete self-sufficiency of individuals. It is logically conceivable that even small transactions costs will shut off all trade, if tastes and endowments are sufficiently similar. But even if tastes and endowments are such that individuals have to trade to satisfy any of their needs, increasing transactions costs can reduce the

---

[16] For numerical illustrations of different monetary arrangements, based on a simplified model, see Niehans (1969).

[17] The square bracket in equation (5) is the marginal gain from the direct sale of one unit of *h* for *k*. The marginal gain from the corresponding indirect exchange with *g* as medium of exchange is

$$p^{hg}\lambda^{g} - \lambda^{h} - \lambda^{q}(c^{h} + c^{g}p^{hg})$$
$$+ p^{hg}[\lambda^{k}p^{gk} - \lambda^{g} - \lambda^{q}(c^{g} + c^{k}p^{gk})]$$

Deducting the first expression from the second, we obtain the marginal gain from using the indirect route

$$(p^{hg}p^{gk} - p^{hk})(\lambda^{k} - c^{k}\lambda^{q}) - 2c^{g}p^{hg}\lambda^{q}$$

The first term represents the value of the gains in *k* from using the more favorable indirect prices, net of additional transfer costs, while the second term measures the value of the additional transactions costs on the medium of exchange. There is clearly some c^{g} high enough to make this negative.

volume of trade to any level we care to specify. Suppose no work is directly consumed, thus maximizing the amount available for transactions. The exchange programs which can be maintained with this amount at the given transactions cost rates specify a "trade possibility frontier." By increasing transactions cost rates, this frontier can be made to shrink, until finally the smallest trace of market exchange would use up all of society's work.[18] By increasing transactions costs we can thus bring society as close to individual self-sufficiency as we want.

Going to the other extreme, we may reduce all transactions costs to zero. In this case, the model reduces to a neoclassical general equilibrium system. Equations (4) and (6) show that for all nonzero flows, prices will now equal the marginal rates of substitution. The price system will be consistent in the sense that the direct exchange rate for any two commodities equals the quotient of their prices in terms of a third good. As a consequence, the resource constraints can be expressed in terms of any one of the goods, arbitrarily chosen as numeraire, and added over all goods. The result is the familiar budget constraint.[19] The price we pay for these simplifications is that exchange flows at the various trading posts become indeterminate; it is irrelevant to what extent apples are traded against nuts, oranges, or gold dust. From the point of view of monetary theory the neoclassical system has often been criticized for being essentially a barter theory. In the light of these considerations this criticism does not seem valid. Its basic shortcoming is rather that the choice between barter and monetary exchange and, in the case of monetary exchange, the choice of the medium (or media) of exchange, are matters of economic indifference. As a consequence, a neoclassical system is unable to give an explicit analysis of monetization and monetary services.

Finally, starting again from the randomly assigned transactions cost rates, we reduce these rates to zero for one commodity, denoted *m*, leaving all others unchanged.[20] The availability of just one good with zero transactions costs is enough to guarantee consistency of the equilibrium price system in the sense defined above.[21] Once consistency is assured,

---

[18] Consider (3), disregarding time subscripts and stocks and letting $x^q = 0$. Adding over individuals

$$\sum_{i}^{n} \sum_{h}^{q} \sum_{k}^{q} (c_i^h + c_i^k p^{hk}) y_i^{hk} = \sum_{i}^{n} \bar{x}_i^q$$

since aggregate market purchases and sales cancel in equilibrium. It is clear that for a given endowment with work the y's can be forced down to any level by increasing the c's.

[19] Consider equations (2) and (3), again disregarding time subscripts and stocks and redefining h to include q (since there are no transactions costs). Multiply the constraint for h by p^{hN}, the price of h in terms of the numeraire, and add over commodities, remembering that $p^{kh} p^{hN} = p^{kN}$. The result is

$$\sum_{h}^{q} p^{hN}(x^h - \bar{x}^h) - \sum_{h}^{q} \sum_{k}^{q} (p^{kN} y^{kh} - p^{hN} y^{hk}) = 0$$

But the double sum is identically zero, whence

$$\sum_{h}^{q} p^{hN}(x^h - \bar{x}^h) = 0$$

[20] If transactions costs are lowered for particular traders, instead of for particular goods, these traders will assume the position of middlemen in the exchange system. For an example see Niehans (1969, pp. 720 ff).

[21] In case this is not evident, it can be proved as follows. Suppose, contrary to the consistency postulate, $p^{hm} p^{mk} < p^{hk}$. If $y^{hk} > 0$, (6) holds as an equality, whence

$$p^{hm} p^{mk} < \frac{\lambda^h + c^h \lambda^q}{\lambda^k - c^k \lambda^q}$$

But for the two legs of the indirect exchange

$$p^{hm} \leqq \frac{\lambda^h + c^h \lambda^q}{\lambda^m - c^m \lambda^q} \quad \text{and} \quad p^{mk} \leqq \frac{\lambda^m + c^m \lambda^q}{\lambda^k - c^k \lambda^q}$$

and thus, for $c^m = 0$,

$$p^{hm} p^{mk} \leqq \frac{\lambda^h + c^h \lambda^q}{\lambda^k - c^k \lambda^q}$$

But we already know that the inequality sign holds in this case, which means that the indirect exchange does not take place. If, on the other hand, we start out by assuming that the indirect exchange takes place, the same type of argument shows that the direct transaction violates equation (6). It follows that with $p^{hm} p^{mk} < p^{hk}$ there is only a direct supply of h against k. In the same way it can be shown that with $p^{hm} p^{mk} > p^{hk}$, only the indirect route will be used. However, market equifurther conclusions can be drawn. First, if a good with zero transactions costs is available, it is inefficient to use any other good as medium of exchange. The commodity m will thus be elevated to the status of the general medium of exchange, called money. With consistent prices, we only have to compare transactions costs in the various markets, and these by assumption are always lower for m. All nonmonetary goods will thus flow directly from the initial resource owner to the final consumer. Second, there will never be an advantage in using direct barter instead of monetary exchange; the nonmonetary trading posts may just as well be closed down. Third, all commodity flows can again be arranged as if the individual were under no constraint to balance his accounts in each market separately; an aggregate budget constraint is all that matters.[22] We are thus close to a neoclassical model, but with a determinate monetary exchange network.[23] It should be noted, however, that even in this case money is still commodity money, being currently supplied and consumed. As a consequence, the money of this section, even if it can be transferred without cost, is not neutral and the quantity theory of money (in the sense of proportionate effects of the money supply on prices) does not apply.

V. Token Money

A timeless flow model is unable to explain the stock demand for money. It is also unable to explain the use of a token money circulating in given quantity with zero flows of new supply and consumption.[24] It thus precludes an analysis of some of the most fundamental aspects of money. To provide such an analysis we have to return to the complete model including the time dimension, stocks and storage costs.[25]

Suppose we start again by assigning positive random values to transactions and storage cost rates, resulting in a pattern of exchange flows and stocks without any particular monetary characteristics. In contrast to the timeless case, certain goods may now be exchanged and held in stock even though they fail to appear in anybody's utility function. Their utility is of a purely indirect nature, being derived from their efficiency as media of exchange. Since they are not consumed, the current endowments are all used up by storage costs. By assigning low enough values to the respective transactions and storage costs, it is always possible to make one of these goods, again designated m, the general medium of exchange. This will then be a token money.[26] The current endow-

---

librium requires that every supply of h against k be matched by a corresponding supply of k against h. Examining equation (6) for flows in this direction, it turns out that for $p^{hm}p^{mk} < p^{hk}$, the supply of k will travel along the indirect route. We are thus confronted with the typical arbitrage situation, where h is supplied in the direct barter market, while the corresponding supply of k takes its way through other markets. The same is true, with the reverse sign, for $p^{hm}p^{mk} > p^{hk}$. It follows that there can be no equilibrium with inconsistent prices.

[22] Identify one good as money, let all nonmonetary transactions in equations (2) and (3) vanish and substitute for the y's from all other constraints into the money constraint. For the timeless case, the result is the budget constraint

$$\sum_{k}^{q} p^k(x^k - \bar{x}^k) + (x^m - \bar{x}^m) + \sum_{k}^{q} p^q c^k \left| x^k - \bar{x}^k \right| = 0$$

with transactions costs as a nonclassical component.

[23] An element of indeterminacy is still left inasmuch as the hungry tailor and the shivering baker, if they should happen to meet, are indifferent between direct barter and two compensating monetary exchanges. In the more general case with nonzero transactions costs on money, this is, of course, not so.

[24] One could construct a timeless flow model where money is supplied in positive amounts without being consumed, all current supplies being used up by transactions costs. This would require, however, that the commodity which is used as money is, at the same time, the one used up in making transactions, thus taking the place of work in the present model. This is clearly a highly artificial assumption.

[25] For illustrative computations of the stock demand for money in a model with predetermined consumption, see Niehans (1970).

[26] If the endowment flows of a token money are posi-

ments with token money cover the phenomenon of "seignorage" for a stationary economy. If there is some attrition of the money stock, those who are able to supply the flow of replacement money obtain a steady stream of other goods and services from those who wear out the money, but do not have the power to replace it.

We now assume that money, in addition to being token money, has zero transactions costs. Inspection of the programming problem shows that the system then becomes homogeneous of degree zero in the given monetary resources ${}^t\bar{x}_i^m$ for all i and t, and in money prices ${}^tp^{km}$ foı all t and k.[27] We thus end up with neutral money. While for each money user at given prices such a money has a positive marginal (flow) value, measured by ${}^t\lambda^m$, a general increase of all monetary endowments in the same proportion will leave everybody's utility unaffected, all ${}^t\lambda^m$ moving inversely to money prices.[28] The absence of direct utility and transactions costs for money thus appear as the crucial assumptions for neutrality. If desired, the system can now be dichotomized in the neoclassical way.

Finally, let us suppose there is a token money for which there are no current endowment flows, that is ${}^t\bar{x}^m = 0$. Correspondingly, holding costs must be zero ($\gamma^m = 0$); otherwise the terminal stock could not equal the initial stock. The economy can now be thought to be equipped with an arbitrary stock of such a "perfect money"

$$\overline{M} = \sum_i^n {}^t s_i^m \qquad \mathrm{t} = 0 \ldots \mathrm{T}$$

This fixes the equilibrium values of absolute money prices and individual cash balances at every moment of time. Since the marginal stock value of money measured by ${}^{t+1}\lambda^m - {}^t\lambda^m$ is now zero (see equation (8)), the economy is satiated with cash balances.[29] This is the highly special limiting case envisaged in the neoclassical approach to monetary equilibrium, but now made determinate by transaction costs on nonmonetary goods.

VI. Concluding Remark

In this paper the neoclassical system was extended to include transactions and storage costs, providing a general equilibrium theory of exchange. Within this framework, various exchange systems, ranging from no exchange to multilateral barter and further on to various forms of monetization, could be constructed by suitable assumptions about transactions and storage costs. Within its confines, the transactions cost approach is able to deal effectively with three basic problems of monetary theory inasmuch as it 1) provides full integration of monetary and value theory, 2) gives an explicit analysis of the services of money, 3) includes the neutral money of the neoclassical theory as a limiting case. The main deficiency of the present model is the neglect of time preference. This makes it difficult, if not impossible, to apply the model to those cases in which the cost of holding assets cannot reasonably be interpreted as a physical shrinkage. An analysis of inter-

---

tive, there must still be some storage costs to make the model stationary over the year. In this case prices will be such that the resulting demand for money stocks is just high enough to make attrition equal to the new supply.

[27] In the resource constraint for m, with zero consumption, a doubling of ${}^t\bar{x}^m$ will be reflected in a doubling of stocks, a doubling of the money prices of other goods ${}^tp^{km}$ and of money flows ${}^ty^{mh}$, while the commodities exchanged for money ${}^ty^{hm}$ are unchanged. In all other resource constraints, the money flows are doubled while the price of money in terms of goods is cut in half, leaving their product unchanged. With unchanged commodity flows and zero transactions costs on money, transactions costs in the constraint for q are not affected either.

[28] Consider equation (6) for the money price of h. For $c^m = 0$, it reads ${}^t\lambda^m {}^tp^{hm} = \lambda^h + \lambda^q c^h$. All that matters is the product of λ^m and ${}^tp^{hm}$ which does not change.

[29] Of course, satiation with cash balances does not mean that the marginal flow value of money ${}^t\lambda^m$ goes to zero—the loss of a dollar is always a loss, even if I am satiated with cash balances.

est, credit, financial intermediation, and the rules for an optimal money supply, will thus require a further extension of the model.

REFERENCES

K. J. Arrow and A. C. Enthoven, "Quasi-Concave Programming," *Econometrica,* Oct. 1961, *29,* 779–800.

M. L. Balinski and W. J. Baumol, "The Dual in Nonlinear Programming and its Economic Interpretation," *Rev. Econ. Stud.,* July 1968, *35,* 237–56.

K. Brunner and A. H. Meltzer, "Some Further Investigations of Demand and Supply Functions for Money," *J. Finance,* May 1964, *19,* 240–83.

J. R. Hicks, "Gleichgewicht und Konjunktur," *Zeitschrift für Nationalökonomie,* 1933, *4,* 441–55.

J. Niehans, "Money in a Static Theory of Optimal Payments Arrangements," *J. Money, Credit, Banking,* Nov. 1969, *1,* 706–26.

———, "Die Geldnachfrage in einer dynamischen Optimierungstheorie des Zahlungssystems," *Schweizerische Zeitschrift für Volkswirtschaft und Statistik,* June 1970, *106,* 129–48.

P. N. Rosenstein-Rodan, "The Coordination of the General Theories of Money and Price," *Economica,* Aug. 1936, *3,* 257–80.

P. A. Samuelson, "What Classical and Neoclassical Monetary Theory Really Was," *Can. J. Econ.,* Feb. 1968, *1,* 1–15.

The Uses of Money: Money in the Theory of an Exchange Economy

By KARL BRUNNER AND ALLAN H. MELTZER*

One of the oldest unresolved problems of monetary theory is to explain the use and holding of money. Resolution of the problem is central to an understanding of the difference between a monetary and a nonmonetary or barter economy, and there have been numerous attempts at resolution. The use of money or the existence of a positive demand has been made to depend on such diverse factors as the anticipation of price or interest rate changes, uncertainty, the embarrassment of default, legal restrictions, or some undefined set of "services"—such as "liquidity"—that money provides. Probably the dominant, current explanation posits the existence of a number of ostensibly separate "motives for holding money." This explanation coexists with vestiges of an earlier preferred explanation that made the use and/or holding of money depend on the "functions" performed by money, such as medium of exchange, store of value and unit of account.

Four recent developments have increased economists' interest in the problem. First, money has remained in use in a number of countries even in periods of accelerating inflation, most recently in Brazil. The continued acceptance of existing monies under conditions of ever-increasing holding cost calls into question the relevance of treating money as an asset that provides little or no return. Second, proponents of a new international monetary unit have claimed benefits to the world economy that exceed the saving of resources arising from the substitution of paper money for commodity money. Third, recent work (see Boris Pesek and Thomas Saving, Milton Friedman and Anna Schwartz) has revived interest in the properties of a medium of exchange, in the appropriate definition of money, and in the relation between the assets that serve as medium of exchange and the assets that serve as money. Fourth, growth theorists have introduced an asset called "money" into models of economic growth. Some of their studies (Miguel Sidrauski, James Tobin (1965)) conclude that the introduction of money either reduces the community's nonmonetary real wealth and lowers the capital-labor ratio or has no effect at all on society's real wealth.[1] The economy of these growth models becomes monetized only because of a difference between private and social returns from the use or holding of money.

A large literature has developed to explain why individuals and/or societies forego consumption so as to hold an asset that adds nothing to individual wealth or utility,[2] or in some versions, has negative

---

* University of Rochester and Carnegie-Mellon University. We remain indebted to the National Science Foundation for their continued support of our work. We have benefitted from the helpful suggestions of Jack Hirshliefer, Thomas Saving, Jerome Stein, Terry Turner, and especially Axel Leijonhufvud. Their many comments and suggestions improved the paper.

[1] Quoting Tobin, ". . . as viewed by the inhabitants of the nation individually, wealth exceeds the tangible capital stock by the size of what we might call the fiduciary issue. This is an illusion, but only one of the many fallacies of composition which are basic to any economy or any society" (1965, p. 656).

[2] This position is very old and was stated clearly by Henry Thornton, one of the best of the early monetary theorists (1965, p. 234) when discussing the holding of Bank of England notes.

> "It presents to the holder no hope of future profit from the detention of it. Not only does it bear no

marginal social product.[3] Aspects of this literature have been surveyed several times (John Gilbert, Arthur Marget, Will Mason, Don Patinkin) and summarized by Keynes (1936, ch. 15) who based three of his four motives for holding money on the principal arguments advanced by his predecessors. The main arguments, used alone or in combination, invoke 1) time, 2) uncertainty, 3) lack of synchronization of receipts and expenditures, 4) costs of transacting, and 5) the existence of nonpecuniary returns. Of these, the most common is one or another of the many versions of the synchronization argument. This argument is valid only in the simple barter economy of the textbooks.[4]

Friedman (1956) and Patinkin recognized that previous analyses yielded scarcely any information about the productivity of money. Building partly on work by Paul Samuelson (1947, pp. 118–20) they identified the productive services of money with a nonobservable, nonpecuniary return to money. But the nature of the services yielded by money and the conditions governing the marginal productivity of money have not been much explored in their work. The most extensive discussion of the coexistence of money, bonds and real capital in portfolios explains the use of money by asserting that, "Intuitively, money seems to be a more efficient carrier of nonpecuniary services . . . than bonds . . ." (Friedman (1969, p. 25)).[5]

The most casual observation suggests that a limited number of assets is used to make or receive payments in all but a few primitive societies. Very similar assets are used in very different places. The standard theory of exchange, or price, however, provides no hint as to why dominant mediums of exchange emerge.[6] Any asset or combination of assets is equally likely to be selected as a medium of exchange,

---

> interest, but it offers no substitute for interest; . . . the quantity held by each person is only that which the amount of payments to be effected by it renders, in his opinion, necessary."

However, in Chapter III, Thornton recognizes that resource costs are reduced when paper money replaces gold and silver and (pp. 91 ff.) he criticizes Smith, whom he credits with this argument, for failing to see that resources are also saved when bills of exchange replace notes.

[3] Recent exceptions are David Cass and Menahem Yaari and Jürg Niehans (1969) who show that social efficiency rises if the community agrees to use money. Cass and Yaari discuss the use of money as a medium of exchange (or in their terminology intermediation) and show that money is productive under the circumstances of their model. However, they restrict barter exchange to a single "double coincidence" transaction and assume that each trader knows, or can obtain at zero cost, information about the exchange ratios or prices at which all other transactors are willing to trade. Our earliest attempt to relate the productive contribution of money to the saving of costs of information was published in 1964 (pp. 258–62).

[4] It is easy to see why "lack of synchronization" does not imply that money is used and held. Consider an economy that has neither a medium of exchange nor money. If there are no costs of acquiring information, differences in the timing of receipts and payments are adjusted by issuing verbal promises in exchange for goods and, later, delivering goods. More generally, in a barter-credit economy, commitments or promises to pay "bridge the gap between receipts and payments." Gilbert discusses several versions of the "synchronization" argument and provides references to the more prominent writers and to those who stated their arguments most clearly. We have found no one who discusses barter-credit. A simple statement of the synchronization argument is due to Joseph Schumpeter (p. 547) who argued that if people are paid on Saturday and stores are closed until Monday, firms hold money from Monday to Saturday and households from Saturday to Monday because "the institutional arrangement so wills it." For a comparison with current notions of the Hicksian week, see Patinkin (pp. 13–14).

[5] Friedman does not explain why money is superior to bonds as a carrier of nonpecuniary services. In his analysis, the only uncertainty is about the timing of receipts. We find nothing to prevent individuals in his model from holding interest yielding bonds and using verbal promises to pay when expenditures temporarily exceed receipts.

[6] Although specialization has long been invoked to explain exchange, economic theory explains specialization in exchange as the result of a corner solution. For much the same reason that theory offers no explanation of the concentration of the medium of exchange function in a narrow range of assets, theory provides no explanation of the specialization of traders in other inventories and types of exchange.

and the allocation of resources is not affected by the choice. In the next two sections we reconsider the services rendered by money, extend the theory of choice to include the choice of the assets individuals use as mediums of exchange and then use diagrams to illustrate the main points of our analysis.

Our argument requires the use of two postulates that are not part of conventional exchange theory. They are:

1) For each transactor in an exchange economy, the marginal cost of acquiring information, measured in units of consumption sacrificed, depends on the goods or services selected.
2) The marginal cost of acquiring information about the properties of any asset does not vary randomly within a social group and declines as the frequency with which the group uses a particular asset increases.

The two postulates are necessary and sufficient conditions for the use of a medium of exchange. They emphasize the point that will concern us throughout—that it is the uneven distribution of information, and not the existence of an undifferentiated uncertainty, that induces individuals to search for, and social groups to accept, alternatives to barter.

Social choice of an asset used as money is separate from, though not entirely independent of, individual decisions to hold money. We believe it is useful to separate the analysis of the choice of the asset used as money from the analysis of the optimal amount of money to produce and from the individual's choice of a desired money balance. Our interest here is in the individual and social choice of the assets used as money, the services money provides to individuals and societies, the relation of these services to the choice of a monetary unit, and some implications of these decisions.

I. The Services of Money to Individual Transactors

One of the main productive activities of a household in a developed market economy is the acquisition of the goods and services consumed by the household. The provision of these goods and services requires not only the sale of income-yielding productive services but the use of resources to acquire information, arrange payments, and schedule purchases. Shopping, budgeting, and planning expenditures are productive tasks that both absorb resources and yield benefits to the skilled or knowledgeable purchasers or sellers who make advantageous exchanges. The use of a medium of exchange permits the household to economize on the amount of resources absorbed by these activities and to enjoy a larger and more diversified basket of goods and more leisure.

Potential transactors possess very incomplete information about the location and identity of other transactors, about the quality of the goods offered or demanded, or about the range of prices at which exchanges can be made. Uncertainty about quality characteristics is a main reason for the dispersion of prices of any commodity, and uneven distribution of information about the qualities of commodities increases the dispersion of prices both within the community and between bid and ask quotations.[7] Transactors can acquire information most readily and at lowest cost about commodities that are most widely used and best known, so the prices of these commodities have the least dispersion. When the qualities of a commodity are less certain, acquisition of information requires greater use of valuable resources, and dispersion increases. If risk aversion is prevalent, uncertainty about the properties or quality of an asset further lowers the average demand price of the asset.

---

[7] Relatively low cost of acquiring information and a smaller dispersion of prices is one of the many meanings of the term liquidity, a property often attributed to money. Reducing the cost of acquiring information raises liquidity. We conjecture from this line of reasoning that the announcement that shares are to be moved from the over-the-counter market to an organized exchange such as the New York Stock Exchange raises the average price of the shares. The price increase is a measure of the value to the stockholders of the reduced dispersion of bid and ask prices.

The use of money reduces uncertainty and expands trade in a number of ways. One way is by providing a unit of account, or standard in which prices are expressed. If there are N commodities in a barter economy and the unit of account is randomly selected, anyone wishing to organize or participate regularly in a market must know (or be able to obtain) each of the $N(N-1)/2$ independent, exchange ratios in the barter exchange matrix. Social choice of a unit of account reduces the matrix to an Nx1 vector of exchange ratios expressed in the unit of account. The cost of acquiring, processing and storing information falls. The gain from the use of a unit of account is analogous to the gain that comes from introducing a common unit of measure such as height, weight or temperature.

The gain from using a unit of account is limited by the size of the market, but the size of the market expands after agreement on a unit of account because resources devoted to trade receive higher net returns. As information about the unit spreads and the market expands, additional private and social benefits result from the development of the market system. Additional exchange ratios are expressed in the unit; it becomes efficient to use the unit where other units were used previously.[8]

A second and considerably more important way in which the use of money reduces uncertainty and contributes to the expansion of trade and the market system is through service as the medium of exchange. The frequency with which the same unit serves both as medium of exchange and as unit of account suggests that it is efficient to perform both functions with a single unit, but the functions are distinct and require separate analysis.

To analyze the medium of exchange function, we consider a transactor who has an initial endowment of resources including his own labor time and some information about exchange ratios and qualities of commodities. He has several alternative ways of transforming his initial endowment into a preferred bundle. As in standard price theory, he can use his endowment for production, consumption, or exchange. In addition, he has two options that are neglected in traditional price theory. 1) He can use resources to increase his information about the qualities of goods and opportunities for exchange. If the transactor uses resources in this way, we say that he invests in information. 2) He can engage in indirect or roundabout methods of exchange, accepting goods with low marginal cost of acquiring information, transferring and storing, then exchanging these goods for others until he obtains an optimal bundle. We describe the resources allocated in this way as the (real) costs of transacting or exchanging. The resources used in this way are, of course, distinct from the resources exchanged.

Under conditions of uncertainty[9] about the quality of goods offered in exchange and about prevailing market opportunities, the costs of acquiring information and exchanging are neither zero nor identical for every good or service. Our first postulate, introduced above, recognizes that the marginal cost of the resources the transactor uses to acquire information or to carry out transactions is the amount of consumption or endowment sacrificed. This marginal cost depends on the goods or services he selects (or about which he chooses to acquire information) and is subject to substantial variance.

---

[8] An example is the spread of the decimal and dollar system or the metric system to countries where the pound or foot has long served as the standard. The Euro-dollar market suggests the way in which a unit of account begins to acquire the properties of a medium of exchange and other attributes of money through the efforts of private traders. Conversely, where trade between countries is small and infrequent, it is not profitable to devote resources to establishing a unit of account or medium of exchange and barter-credit predominates.

[9] Note that the uncertainty in our account differs

By choosing a sequence of transactions—a transaction chain—involving assets with low marginal cost of information, the transactor can lower the marginal cost of exchanging. He incurs transfer and carrying costs and uses existing information about the qualities of particular goods instead of investing resources to acquire information about other goods or other trading arrangements. However, transfer costs increase with the length of the transaction chain, forcing the rational transactor to compare the marginal cost of acquiring information to the marginal cost of rearranging the transaction chain and to the benefits obtained from these and alternative uses of resources.

The formal statement of our analysis involves the maximization of utility subject to a budget constraint under conditions of uncertainty about market opportunities and the qualities of goods. An individual seeks to obtain the optimum combination of goods and services by investing in information and engaging in exchange. Numerous sequences of transactions are open to him. His problem is to find the optimal sequence of transactions and the optimal investment in information while choosing an optimal bundle of goods or consumption plan. Production of the standard type is disregarded.

Let the individual's utility function be

$$(1)\ U = U(\epsilon, v; Q)\ ; U_1 > 0; U_2 < 0; U_3 > 0$$

where ϵ is the expected bundle of goods the transactor obtains by allocating resources to the various options, v is the variability of the bundle associated with the uncertainty about market conditions, and Q is the information available to the transactor about the qualities or properties of goods. We regard Q as a partition of the state space summarizing the full range of qualities. A skilled trader or market professional has valuable information about the qualities of goods, the location of other traders and their tastes and preferences. His market opportunities differ from those of an unskilled trader or infrequent participant in the market. The more certain a transactor is about the qualities (or properties) of goods, and the more he is able to discern differences in performance or other quality characteristics, the higher his utility. The more information the transactor has about market conditions and the characteristics of goods, the lower the variability of the bundle he obtains and the higher his utility.

The budget constraint is

$$(2)\qquad R_0 = C_0 + X_0 + I_0 + S_0$$

The transactor can allocate his resources, R,[10] to reservation demand, C, to exchange,

---

from the types of uncertainty emphasized in past discussions of money. The latter include uncertainty about 1) the price level, 2) the timing of receipts, and 3) interest rates (see Gilbert). For other recent attempts to relate search and costs of acquiring information to resource allocation, see Herbert Simon and George Stigler.

[10] The variables R_0, C_0, X_0, I_0, and S_0 denote diagonal nonnegative matrices. Each of the symbols is defined in the text, but some discussion of the difference between allocations to S_0 and I_0 may clarify the analysis that follows. No "generally accepted" medium of exchange is imposed on the economy discussed in the text. A transactor can forego consumption and use resources to sample the exchange ratios at which other transactors (including potential transactors) are willing to offer goods of the quality he desires for the goods he presently holds. We use the symbol I_0 to describe the resources

X, to acquire information about market opportunities, I, or to execute transactions, S. The conditions for advantageous exchange depend on the information available to the transactor. By using resources to acquire information, he lowers the amount of resources remaining available for reservation demand and trade. However, the new information reduces uncertainty about market conditions and exchange opportunities and lowers the variability of the return from trading. Optimal information is achieved when the marginal utility of resources withdrawn from consumption (reservation demand) equals the marginal utility of an improved trading position and reduced uncertainty.

Transactors' uncertainty with respect to market opportunities is conditioned by a distribution of exchange ratios governing the social exchange process.

$$\pi = \pi(E \mid P) \tag{3}$$

The term E is a matrix of exchange ratios whose generic elements e_{ij} denote the jth unit obtained per unit of i, and P summarizes the information about market opportunities and exchange ratios available to transactors.[11] Since bid and ask prices are generally not the same in the presence of costs of acquiring and maintaining information and transferring goods, the matrix is not anti-symmetric, and $e_{ij}e_{ji} \neq 1$.

At any time the transactor knows a maximal E, denoted E^*, that describes the best market opportunities available to him if he relies on the information carried over from the past. We treat E^* as a random variable conditional on the transactors' information about market opportunities, P. There is a unique expected value of E^* corresponding to each state of knowledge about market opportunities. A transactor can improve his information about market opportunities by using a portion of his endowment to acquire additional information, i.e., by investment in information. Although a larger sample does not necessarily increase the maximal sample value, under weak constraints on π, a larger sample raises the expected value of the sample maximum and reduces the variance of the exchange ratios at which the transactor trades. We postulate that the mean and variance of the distribution, π, depend on the amount invested in information and the transaction arrangements, T and are denoted

$$\epsilon(E^* \mid I, T) \text{ and } v(E^* \mid I, T) \text{ with}$$

$$\frac{\partial \epsilon}{\partial I} > 0; \quad \frac{\partial v}{\partial I} < 0; \quad \frac{\partial^2 \epsilon}{\partial I^2} < 0; \quad \frac{\partial^2 v}{\partial I^2} > 0$$

Once a transactor allocates I_0 to improve

---

allocated to sampling or search of this kind. The goods the transactor holds may include "intermediate" goods, goods acquired in previous transactions for use in later transactions. Alternatively, a transactor can seek to improve his knowledge of transaction arrangements by 1) learning about the properties of goods that other transactors desire to acquire either for consumption or as intermediate goods, or 2) by learning about the relative costs of storing or transporting goods, or 3) by learning about the costs of completing exchanges using different sets of intermediate goods. We use the symbol S_0 to summarize the resources used to acquire information about transaction arrangements and to carry out exchanges. Resources must be used for similar purposes even if a few assets are generally accepted as means of payment. "Shopping" is a common means of acquiring information about price-quality combinations. Resources must also be used to learn about the advantages of immediate payment over deferred payment, the costs and benefits of using coin or currency rather than checks in general and in specific exchanges, etc.

[11] P is a vector whose coordinates summarize market information. The coordinates depend on the individual's and the community's information. Increased investment in information refines the available information, and reduces the generalized variance. Technological change reduces information and increases the variance of π by reducing information about the qualities of goods. The coordinates of P are the partitions, Q, of the state spaces—characterizing each individual's information. An example helps to describe Q and P. An individual's utility increases if, with unchanged market opportunities, he acquires information that permits him to develop his tastes. An ability to distinguish between Botticelli's and "paintings" increases Q but does not change other elements of P. If everyone acquires the same discriminating tastes, all Q's change and therefore P changes.

his information, only $R_0 - I_0$ of the initial endowment remains available for consumption and exchange. Exchange operations involve allocation to two distinct components, X and S, with X the bundle offered in exchange and S the resources used to produce the exchange. An exchange matrix T[12] characterizes the transformation via exchange transactions of a bundle X_0 into a new bundle, Y_1. The off-diagonal elements of the T matrix are the proportions of the ith good exchanged for the jth good. The resource cost (S) of executing a set of transactions T depends on T; $S = S(T)$. A market exchange is a transformation of the inputs X, $S(T)$ to an output $\epsilon(Y_1)$. The selection of the amounts offered in exchange, X, and the division of resources between C, X, and T depends, of course, on the information summarized by $\epsilon(E^*)$ and $v(E^*)$ and therefore on I and T.

The total volume of physical resources available to the transactor after the market transformation is expressed by the diagonal matrix $\epsilon(R_1)$[13]

$$\epsilon(R_1) = \epsilon(Y_1) + C_0 \tag{4}$$

and consists of the reservation demand C_0 plus the bundle $\epsilon(Y_1)$ the transactor expects to obtain via market exchange. Replacing X_0 in the description of $\epsilon(Y_1)$ (from fn. 13) we can express $\epsilon(Y_1)$ as a function

$$\begin{aligned} \epsilon(Y_1) &= Y_1[R_0 - C_0 - I_0 - S(T_0), I_0, T_0; \pi] \\ &= \sum_j X^j[R_0 - C_0 - I_0 - S(T_0), T_0] \\ &\quad \cdot \epsilon(E_j^*(I_0, T) \end{aligned} \tag{5}$$

The function $\epsilon(Y_1)$ depends on the initial endowment R_0, the reservation demand C_0, the transaction arrangement T_0, the investment in information I_0, the distribution of exchange ratios and, therefore, on E^* via π. The mean, ϵ, and variance, v, of the transactor's utility function can now be expressed as the functions

$$\epsilon = \epsilon[Y_1 \mid R_0 - C_0 - I_0 - S(T_0), T_0, I_0] + C_0 \tag{6}$$

$$v = v[Y_1 \mid R_0 - C_0 - I_0 - S(T_0), T_0, I_0] \tag{7}$$

A transactor can change his utility by three distinct types of allocation, represented in our analysis by C_0, I_0, and T_0. The transactor trades whenever the utility associated with $\epsilon = C_0 = R_0$ and $v = 0$ is less than the utility associated with some $C_0 < R_0$.

Taking derivatives of ϵ and v with respect to I_0 and solving the resulting equations shows that the optimal investment in information depends on the effects on mean, ϵ, and variance, v. Increased investment (I_0) either lowers reservation demand (C_0) or the resources devoted to exchange (S_0, X_0). With S_0 and X_0 held constant, an increase in I_0 lowers the variance $v(Y_1)$ and has two, opposing effects on the mean $\epsilon(R_1)$. Increased I_0 raises, and the associated reduction in reservation demand lowers $\epsilon(R_1)$. The investment in information raises the expected maximal ratio, expressed by $\epsilon(E^*)$. The rise in $\epsilon(E^*)$

---

[12] T is a nonnegative hollow matrix whose element t_{ij} describes the portion of the ith good allotted in exchange for the jth good and whose diagonal elements, $t_{ii} = 0$. Also, $1 \geq t_{ij} \geq 0$ for $i \neq j$, and $\sum_j t_{ij} = 1$.

[13] The trade matrix $T'X$ can be described as a sum of matrices X^j, i.e., $T'X = \sum_j X^j$. This decomposition is uniquely determined by the form of the matrix $T'X$. The generic matrix X^j consists of zero row vectors with the exception of the jth row vector which is identical to the jth row vector of $T'X$. The index j ranges over all the rows of $T'X$. Under the circumstances specified, the constituent matrices X^j are functions of the exchange X and the transaction T, i.e., $X^j = X^j(T, X)$. Once the decomposition of $T'X$ has been defined a matrix $\epsilon(E_j^*)$ is associated with X^j. The matrix $\epsilon(E_j^*)$ consists of zero column vectors with the exception of the jth column which is identical to the jth column of the matrix $\epsilon(E^*)$. The product $X^j \cdot \epsilon(E_j^*)$ denotes a diagonal matrix with zero diagonal elements except in the jth row. This diagonal matrix describes the acquisition of the jth good by means of an exchange X and a transaction arrangement T. The sum $\sum X^j \epsilon(E_j^*)$ is a diagonal matrix describing the result of the market transformation defined by the pair (X, T) and, since E^* is a random variable, the expected value of this sum is $\sum X^j \epsilon(E_j^*)$. The result is denoted by $\epsilon(Y_1)$, i.e.,

$$\epsilon(Y_1) = \sum_j X_0^j(T_0, X_0) \cdot \epsilon[E_{j0}^*(I_0)]$$

changes the optimal amount offered in exchange, X_0, and this in turn changes C_0 and T_0 and the transactor's utility.

The transactor acquires information until the gain in utility from an increment ΔI_0 equals the loss resulting from the decrease in resources available for reservation demand and exchange imposed by the constraint

$$(8) \qquad \Delta I_0 = -(\Delta C_0 + \Delta S_0 + \Delta X_0)$$

The second derivatives of ϵ and v with respect to I_0 assure that the marginal utility of investment in information decreases as investment increases. Moreover, the continued decrease of C_0 and X_0 generates an increasing loss of welfare, and the continued decrease of S_0 imposes increasingly severe constraints on the choice of transaction arrangements and (given $\partial^2 U/\partial\epsilon^2$ to form a negative definite Hessian matrix) a nondecreasing loss of utility. The allocation of resources to acquire information necessarily reaches an optimum.

A similar argument applies to the choice of transaction arrangement or reservation demand. Utility increases as more resources are allocated to market transformations, T, and the wider search for transaction arrangements, S_0. The larger S_0 and $S(T_0)$, the larger the admissible set of transaction arrangements and the greater is a transactor's opportunity to exploit differences in the marginal cost of information, associated with various goods. For example, by using indirect or roundabout exchange arrangements that reduce uncertainty, the transactor can lower v by allocating resources to S_0 while keeping I_0 constant. The increase in S_0 (holding I_0 constant) reduces C_0 and X_0, and therefore lowers ϵ. But the allocation to S_0 permits the transactor to choose from a wider range of transaction arrangements, raising ϵ. The optimal allocation to S is reached when the gain in utility from the smaller variance and more advantageous transaction arrangements just equals the loss in utility due to a smaller C_0 and X_0.

By neglecting costs of search and exchange, the traditional analysis of barter exchange sets S and I to zero and thus omits important aspects of resource allocation and choice under uncertainty. In the usual analysis, a transactor either engages in a single, double-coincidence transaction, $C_1 = R_1$, or does not trade, $C_0 = R_0$. In our analysis, a transactor is not forced to choose between autarchy and a single, double-coincidence transaction. He can engage in a sequence of transactions, expressed by a sequence of matrices or transaction chain $[T_0, T_1, \ldots, T_n]$ chosen so as to exploit differences in the marginal cost of acquiring information and exchanging. The length of the transaction chain—the optimal number of exchanges—is determined jointly with the choice of commodity bundle $C_n = R_n$ that maximizes utility. The bundle, R_n, is defined recursively in equation (9); for convenience the expected value operator is omitted in

---

$$\begin{aligned}
(9) \quad R_n &= Y_n + C_{n-1} \\
&= Y_n + R_{n-1} - [X_{n-1} + I_{n-1} + S_{n-1}(T_{n-1})] \\
&= Y_n + Y_{n-1} + C_{n-2} - [X_{n-1} + I_{n-1} + S_{n-1}(T_{n-1})] \\
&= Y_n + Y_{n-1} + R_{n-2} - [X_{n-1} + I_{n-1} + S_{n-1}(T_{n-1})] - [X_{n-2} + I_{n-2} + S_{n-2}(T_{n-2})]
\end{aligned}$$

$$R_n = \sum_{i=1}^{n} Y_i + R_0 - \sum_{i=1}^{n-1} [X_i + I_i + S_i(T_i)]$$

the statement of the definition. Moreover, every Y_i is a function of T_{i-1}, X_{i-1} and the sequence of investments in information $(I_0 . . . I_{i-1})$ given by

$$(10) \qquad \epsilon(Y_i) = \sum X^j_{i-1}(T_{i-1}, X_{i-1}) \cdot [E_j^*,_{i-1}(I_0, I_1, \ldots, I_{i-1})]$$

as shown in footnote 13. Both ϵ_n and v_n are, therefore, functions of the transaction chain $(T_0, \ldots, T_{n-1})$, the sequence of exchanges $(X_0, \ldots, X_{n-1})$ and the sequence of investments in information $(I_0, \ldots, I_{n-1})$ according to the definition

$$(11) \qquad \epsilon_n = \epsilon(R_n); \qquad v_n = v(R_n)$$

Where information about the qualities and exchange ratios varies with the commodities, there are corresponding differences in the marginal cost of acquiring information, so transactors are not indifferent about the commodities they accept for use in subsequent exchanges. By choosing transaction arrangements T_i that induce the transactor to include in his transaction chain commodities or assets whose qualities are better known and less uncertain, the transactor can reduce the variance $v(Y_i)$ and the resources I_i that he uses to acquire information about exchange opportunities. At each step in the sequence, more resources become available for exchange and for reservation demand, so utility increases. By careful choice of the elements in the triplet C, T, I, a transactor is able to obtain the commodity bundle with mean and variance ϵ_n and v_n that maximizes utility.

Our first postulate, introduced earlier, makes the marginal cost of acquiring information depend on the good or service selected. Unless transactors ignore the information they acquire, the postulate implies that intermediate transaction matrices in the chains they select (i.e., matrices up to the last) do not exhibit a random distribution of nonzero elements. In the first matrix of the chain, clusters of nonzero elements appear most frequently in the columns identified with assets that reduce uncertainty, provide more reliable information and lower marginal information costs. In the last matrix of the chains, the nonzero elements tend to be in the rows associated with commodities that have these same properties.

The distribution of costs of acquiring and maintaining information gives the transactor an opportunity to reduce the resource cost of acquiring his preferred commodity bundle by substituting knowledge about transaction arrangements for investment in information about market conditions and the qualities of goods offered in exchange. Cost reduction occurs in two ways. First, detailed information about market conditions such as location and identification of transactors, the quality and type of commodity bundles they hold and the exchange ratios at which they trade probably decays more rapidly than knowledge about optimal transaction chains. Second as the use of an asset in exchange increases, the transactor learns more about the asset's properties. With growing use of particular transaction chains and improved knowledge of the properties of the assets exchanged, uncertainty and the variances and covariances in the general covariance matrix describing the overall density $\pi(E)$ decline.

The marginal productivity of transaction arrangements can be expressed in terms of the smaller investiment in information required to keep the variance $v(R_n)$ unchanged. By choosing transaction arrangements that reduce the cost of acquiring information and exchanging, a transactor can allocate additional resources to exchange or reservation demand without raising the variance of the bundle he obtains. Indeed, if this were not so, direct exchange would predominate.

The usual discussion of a barter economy neglects the marginal productivity of transaction arrangements by limiting exchange to a single, double coincidence transaction. Once the transactor is given a choice of I and S, both the length of his optimal chain and the assets that enter depend on differences in the productivity and cost associated with alternative chains. Thus, our first postulate assures that the choice of transaction chain and of the assets used in exchange is neither random nor determined solely by the exchange—i.e., by the initial and terminal bundles.

To show that individuals' optimizing behavior leads to the social choice of a small number of medium of exchange assets, we use our second postulate. This postulate states 1) that the marginal cost of acquiring information does not vary randomly within a social group and 2) that the marginal cost declines as the frequency with which an asset is used increases. Transactors can acquire information about a particular subset of the available assets at comparatively low marginal cost once these assets are used frequently.[14]

Our second postulate implies that the transaction chains of the numerous participants in the market process exhibit some common properties. These properties can be expressed as a clustering of the nonzero elements in the columns of the first matrix and in the rows of the last matrix of the chain. The repetitive use of a relatively small number of transaction chains by the members of the social group further reduces the marginal cost to each transactor of acquiring information about the assets most frequently used. The lower costs of acquiring information and transacting induce further clustering and the convergence of individuals' chains toward a common pattern.

There are many stages of development between double-coincidence barter and a fully monetary economy. At some stage, a few assets are used with dominant frequency in transactions. Once this stage is reached, a majority of the transaction chains consists of two matrices, the first containing a few (nonzero) columns and the second a few (nonzero) rows. Thus, money as a medium of exchange, as a transaction dominating asset, results from the opportunities offered by the distribution of incomplete information and the search by potential transactors to develop transaction chains that save resources.

The analysis also explains the emergence of specialized trading functions such as brokerage and other market arrangements. They develop from the conditions that shape the (social) convergence to a dominant medium of exchange. Where information is complete and both information and readjustment are costless, specialization of trading functions yields no economic advantages and has no utility. Where information and readjustment are not free, the situation changes. Specialized services lower the costs of acquiring information and trading by providing more complete information about the range of qualities and market conditions. With a smaller investment of resources a transactor acquires the same information, and more resources can be used for reservation demand or trade.

II. A Diagrammatic Exposition

By simplifying the argument and omitting several aspects, we can use diagrams to demonstrate some main points of our analysis. We start with a nonmonetary economy in which there are two goods. One is the initial endowment of a transactor, R_1; the other is the good, Y_2, acquired by trading in the market. We disregard the re-

---

[14] Differences in technological properties of commodities contribute to differences in the costs of acquiring information and transferring. Monetary literature has long recognized that portability, divisibility, durability, etc. are useful properties for assets used as mediums of exchange.

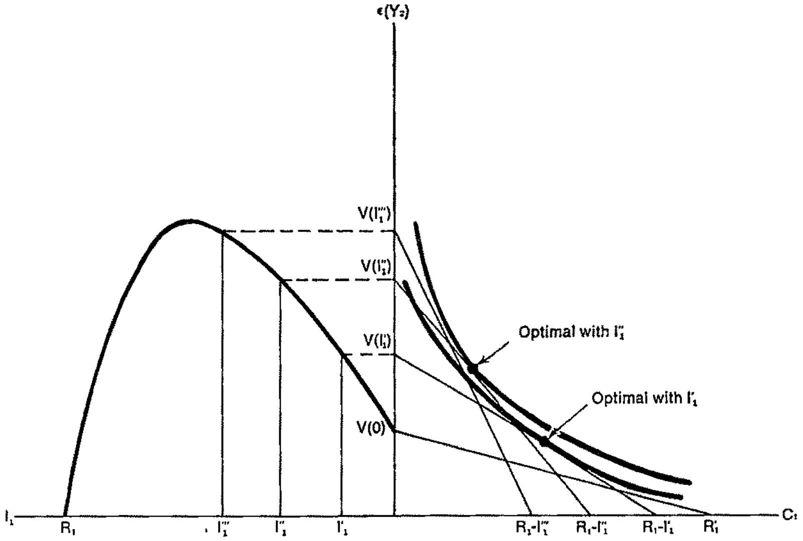


FIGURE 1

sources used in the trading process and allocate the initial endowment to reservation demand, C, to exchange in the market, X, and to the acquisition of information about market opportunities and exchange ratios, I.

$$R_1 = C_1 + X_1 + I_1 \tag{12}$$

The exchange ratio between R_1 (or C_1) and Y_2, denoted e_{12}, states the units of Y_2 obtainable per unit of R_1. Market opportunities are not known with certainty. By investing resources I_1 to acquire information, a transactor obtains a sample of potential exchange ratios. The sample has a maximal value e_{12}^*, the latter a random variable dependent on the underlying distribution. As above, we make the expectation and variance of e_{12}^* a function of investment, I.[15] Market exchange is a productive process that transforms the inputs (X_1, I_1) into an output $Y_2 = X_1 e_{12}^*(I_1)$. The value of the bundle after an exchange is $R_2 = Y_2 + C_1 e_{12}^*$.

The transactor seeks to maximize the utility of the bundle available for consumption. The arguments of the utility function are C_1 and the expected value and variance of Y_2. The indifference curves of Figure 1 are drawn for a given value of the variance $v(Y_2)$. A change in variance moves the entire surface, or what is the same thing, requires a relabelling of the indifference curves to associate a higher level of utility with an unchanged combination of Y_2 and C_1, if the variance of Y_2 decreases and a lower level if utility of the variance increases.

Unlike the standard analysis, neither the slope nor the position of the budget line is independent of the transactor's allocation of resources. Increased investment in information raises the expectation of the maximal sample value e_{12}^* and the slope of the budget line, $-e_{12}^*$. With increased investment, the budget line becomes steeper, and the transactor obtains more Y per unit of C.

The relation between investment in information, the value of information, and the position of the budget line is developed in the left panel of Figure 1. On the horizontal axis, we measure the resources used to acquire information on a scale running from zero at the origin to R_1, the transactor's endowment. By sacrificing a unit of real consumption and investing in information, a transactor enlarges his sample of exchange ratios and, in general, raises the value of e^*. The marginal cost of information is the consumption sacrificed. The value of the information is the gain or loss in $\epsilon(Y_2)$ obtained from the exchange of a smaller (R_1-I_1) endowment of real resources at a higher exchange ratio. The curve in this panel shows the value of information $V(I_1)$ as a function of the amount invested. The intercept of the curve on the Y_2 axis is the value of the information, in $\epsilon(Y_2)$ units, that the transactor carries over from previous trades. If he relies entirely on past experience, the best he can do is trade at the ratio $e^*_{12}(0)$; the maximum Y_2 he can obtain by exchanging all his endowment at this ratio is shown as $V(0)$. By investing in information, at a cost equal to I_1, he increases his information about market conditions. The change in the value of information is shown by the projections of the curve in the left panel on the Y_2 axis. If he invests an amount I_1', he has available for exchange only R_1-I_1', but he can exchange at the higher ratio $e^*_{12}(I_1')$ and obtain a maximum value of $Y_2=V(I_1')$. The distance $V(I_1')-V(0)$ on the Y_2 axis is the value of the information obtained with an investment I_1'.[16]

---

[15] To simplify notation we write e_{12}^* for the expectation of the maximal sample value, and where there is no change of confusion, we use Y_2 instead of $\epsilon(Y_2)$.

For any investment I_1, the value of information is given by

$$(13) \qquad V(I_1) = (R_1 - I_1)e^*_{12}(I_1)$$

Inspection makes clear that if the entire endowment is used to acquire information, nothing is left for trade or consumption, so $V(R_1)=0$. To find the value of I_1 that maximizes $V(I_1)$, we take the derivative

$$(14) \quad \frac{\partial V}{\partial I_1} = e^*_{12}(I_1)\left[\frac{R_1 - I_1}{I_1}\eta(e^*_{12}, I_1) - 1\right]$$

where $\eta(e^*_{12}, I_1)$ is the elasticity of e^*_{12} with respect to I_1. It follows from diminishing marginal productivity of investment that the increase in the expected sample value falls as investment increases, and the elasticity falls as I_1 increases. Moreover, as I_1 approaches R_1, $(R_1-I_1)/I_1$ approaches zero. The value of information V reaches a maximum when the share of the endowment invested in information is

$$(15) \qquad \frac{I_1}{R_1} = \frac{\eta}{1+\eta} < 1$$

If the elasticity is low, investment in information raises the expected sample value and improves market opportunities very little. The maximum V then occurs at a very low value of I_1.

The optimal amount of resources invested in information is determined jointly with the optimal reservation demand and allocation to trade by maximizing utility. Differentiating

$$(16) \qquad \begin{gathered} U = U[\epsilon(Y_2), C_1, v(Y_2)] \\ U_1, U_2 > 0; \qquad U_3 < 0 \end{gathered}$$

with respect to C_1 and I_1, we obtain the necessary conditions for a maximum,

$$(17) \qquad U_1 e^*(I_1) = U_2$$

$$(18) \qquad \frac{\partial V}{\partial I} = C_1\frac{\partial e^*_{12}}{\partial I} + \frac{-U_3}{U_1}\frac{\partial v}{\partial I}$$

---

[16] The cost of acquiring information, I_1, is clearly visible in two places on Figure 1. One is the horizontal axis on the left panel. The other is the difference between R_1 and the intercept of any budget constraint on the C_1 axis, R_1-I_1. The difference defines the maximum C_1 associated with any given I_1. As the figure is drawn, each increment to investment $(I_1''-I_1', I_1'-0)$ and each distance between intercepts of the budget line on the C_1 axis is the same size. The decreasing incremental value of information is shown by the decrease in the distance between the successive intercepts of the budget line on the Y_2 axis. The relations of I to $V(I)$ depends on the vector, P, discussed in fn. 11.

Solving these two equations for C_1 and I_1 and substituting the solutions in the budget constraint $R_1 = C_1 + I_1 + X_1$ gives the optimal allocation to reservation demand, investment in information and trade.

The first of the two necessary conditions is the familiar condition for tangency between the indifference curve and the budget line obtained in the standard analysis of consumer choice. The second equation goes beyond the standard analysis and states the condition for the optimal *position* of the budget line. The equation has two components. The first is the increment to the value of reservation demand, measured in units of Y_2, obtained from investment in information; the second is the marginal gain or loss from reducing the variance. The position of the budget line is optimal when the sum of these components equals the marginal value of information, $\partial V/\partial I_1$, measured in units of Y_2.[17]

The conditions for an optimum are shown in Figure 1. The slopes of the budget lines show the exchange ratio e_{12} for different values of I. The position of the budget lines is determined by the amount invested and the value of the information obtained. The relation between the two is given by the curve in the left panel. We disregard, for the moment, the gain or loss in utility resulting from changes in the variance. As the figure is drawn, it is suboptimal to restrict investment in information to I_1'. The transactor can reach a higher indifference curve by increasing investment to I_1''. A further increase to I_1''' lowers utility. In the absence of any effect of variance, the transacter would choose an investment I_1'' and trad at the corresponding exchange ratio. Th investment I_1'' and the associated alloca tions to C_1 and Y_2 represent an optimun for any transactor unconcerned by th variance, i.e., for any risk neutral transac tor.

For a risk neutral transactor $U_3 \equiv 0$. Fo such transactors, the optimal position c the budget line is always below the prc jection of the maximum value of informa tion on the $\epsilon(Y_2)$ axis. The same is true fo transactors who are not risk neutral when ever $\partial v/\partial I = 0$. In these cases, the optima value of information is always reache before the maximum value of information This can be verified by noting that i these cases $\partial V/\partial I$, therefore $\partial e_{12}^*/\partial I$, mus be positive since $\partial V/\partial I = C \cdot \partial e_{12}^*/\partial I$.

Transactors who are willing to sacrific resources to reduce uncertainty and ob tain more reliable information are impor tant in our analysis. For them, invest ments in information not only shift th budget line but also change the utilit surface. A reduction in variance mean that there is less uncertainty about th outcome of an exchange. With uncertaint reduced, the transactor receives mor utility from a smaller bundle of commod ities. The effect of reducing uncertainty can be shown in Figure 1 by relabellin each of the indifference curves so as t attach more utility to every commodit bundle or by shifting each indifferenc curve in Figure 1 vertically downwar thereby raising the utility assigned to an Y_2, C_1 combination.

The size of the shift in the indifferenc curves resulting from the effect of invest ment in information on the variance pro vides another interpretation of the secon component of the optimality condition The vertical shift, measured in units o Y_2, $(-U_3/U_1)$ $(\partial v/\partial I_1)$, is the amount o $\epsilon(Y_2)$ that provides just as much utility a the reduced variance $v(Y_2)$.

The effect of a reduction in variance o

---

[17] Interpretation of the second component of the second necessary condition is made clearer by noting that $-U_3/U_1 = dY_2/dv$ so the component can be written $[dY_2/dv][dv/dI] = dY_2/dI$. The left side of this last expression shows the change in Y_2 associated with an increment to investment as the product of the change in variance resulting from an increment of information weighted by the appropriate marginal rate of substitution.

utility implies that the optimal position involves more investment in information than I_1''. The bundle available for exchange and consumption declines. In terms of Figure 1, the downward shift in the indifference curves—the gain in utility from reduced variance—at first offsets the loss of utility from the decline in the bundle. Where the marginal reduction in variance $\partial v/\partial I$ and product of the marginal rate of substitution between variance and Y_2, $-U_3/U_1$, is large, the optimal value of information may not be reached until we are on the negatively sloped segment of the $V(I)$ curve. Pronounced risk aversion, a relatively large $-U_3$ raises the portion of the resources a transactor allocates to investment in information and reduction of uncertainty.

We now introduce a third good into the nonmonetary economy we have considered. This good has zero direct marginal utility for each transactor.[18] It contributes to utility indirectly, however, by improving a transactor's information and by reducing the variance of exchange ratios and thus the uncertainty about the bundles obtained in market exchange. We assume that the third good has the following properties:

$$\begin{aligned} &\frac{\partial \overset{*}{e}_{13}}{\partial I} > \frac{\partial \overset{*}{e}_{12}}{\partial I} < \frac{\partial \overset{*}{e}_{32}}{\partial I} \\ (19)\quad &\frac{\partial v(\overset{*}{e}_{13} \mid I)}{\partial I} < \frac{\partial v(\overset{*}{e}_{12} \mid I)}{\partial I} > \frac{\partial v(\overset{*}{e}_{32} \mid I)}{\partial I} \end{aligned}$$

These conditions state that the third good reduces the transactor's cost of acquiring reliable information. With the same investment in information, he increases the expected value of the sample maxima ($\overset{*}{e}_{13}$, $\overset{*}{e}_{32}$) and reduces uncertainty.

The introduction of a roundabout method of exchange using a good that gives transactors increased value of information at relatively low cost offers an opportunity not previously available. Transactors are no longer restricted to choosing between allocations to reservation demand, trade, and investment in information. They may now choose transaction arrangements, i.e., they can obtain Y_2 either by exchanging part of R_1 directly at the exchange ratio $\overset{*}{e}_{12}$ or by first acquiring the new good at exchange ratio $\overset{*}{e}_{13}$ and using good 3 to obtain Y_2 at the exchange ratio $\overset{*}{e}_{32}$.

As before the initial constraint in terms of R is

$$(12)\qquad R_1 = C_1 + X_1 + I_1$$

The transactor wishes to acquire Y_2 by trading X_1 and investing I_1. The market transformation is expressed by $Y_2 = X_1 t^i(I_1)$ where

$$(20)\qquad \begin{aligned} t^1(I_1) &= \overset{*}{e}_{12}(I_1) \\ t^2(I_1) &= \overset{*}{e}_{13}(I_1) \cdot \overset{*}{e}_{32}(I_1) \end{aligned}$$

The transactor desires to maximize the utility function

$$(16)\qquad U[\epsilon(Y_2), C_1, v(Y_2)]$$

by suitable choices of C_1, I_1 and market transformation t^i, subject to the endowment constraint. The optimizing problem can be rewritten as the maximization of the function

$$U[(R_1 - C_1 - I_1)t^i, C_1, v[(R_1 - C_1 - I_1)t^i]]$$

with respect to C_1, I_1 and t^i ($i = 1, 2$). The first-order conditions for C_1 and I_1 are

$$(21)\qquad U_1 t^i = U_2$$

$$(22)\qquad \frac{\partial V(t^i)}{\partial I_1} = C_1 \frac{\partial t^i}{\partial I_1} + \frac{-U_3}{U_1} \frac{\partial v(t^i)}{\partial I_1}$$

---

[18] We do not make this assumption in the formal analysis of the previous section. The assumption is introduced here only to simplify the graphical presentation, specifically to avoid superimposing a second indifference map on the map in the figure. In the more general analysis of the previous section there are a large number of commodities and potential transaction chains. The number of chains depends on the number of commodities. Under our maximizing procedure, at least one of these chains is optimal. In the graphical analysis, V_2 represents this chain.

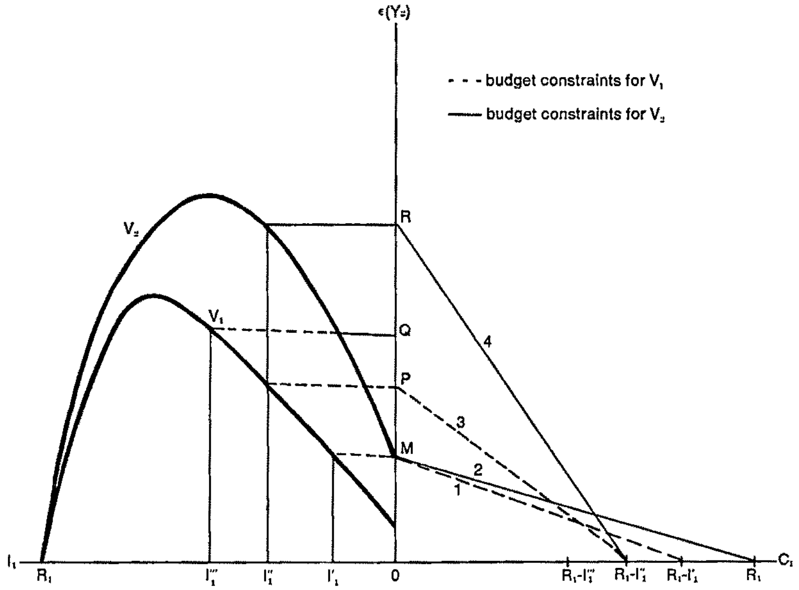


FIGURE 2

where $V(t^i)=(R_1-I_1)t^i$, and $v(t^i)=v[X_1\cdot t^i]$. The two equations determine C_1 and I_1 as a function of t^i. Replacement in the utility function gives

$$U[X_1(t^i)t^i, C_1(t^i), v[X_1(t^i)t^i]]$$

and the stage is set for the choice of the optimal transaction arrangement.

Figure 2 shows the choice of transaction arrangement. With this choice, the transactor determines the value of information obtained from a given investment, I_1. The curves V_1 or V_2 represent two alternative transaction arrangements. The differences in the value of the information obtained from a given investment depends on the properties of the joint conditional distribution $\pi(e^*_{12}, e^*_{13}, e^*_{32}/I_1)$ of the three exchange ratios. Since information about types of transaction chains decays slowly, the intercept of the curve V_2, describing chain t^2, is above the intercept of V_1. The diagram also shows an earlier and higher maximum for transaction chain t^2. Since V_2 lies above V_1 throughout the ascending branch of V_1, a trader gains from using transaction arrangement t^2 throughout this range. This effect is reenforced by the shift of the indifference map resulting from the greater reduction in variance per unit of I. The selection of t^2 and the use of an asset with well-known properties permits the transactor to trade at a more favorable exchange ratio and reduces uncertainty. Thus, the choice of t^2 instead of t^1 raises the transactor's welfare.

Figure 2 shows the productivity gain resulting from the choice of transaction arrangements in two distinct ways. One measure of the gain is the smaller investment in information required to obtain information of given value. For example, the transactor can obtain information with value OM either by investing I_1' and using transaction arrangement V_1 or by making no investment in information and using transaction arrangement V_2. The productivity gain is equal to the saving of resources (I_1'). This saving permits the transactor to trade at the maximal exchange ratio represented by (the solid) budget line 2 instead of the maximal ratio represented by (the dashed) line 1. Similarly, point OQ shows that the combination of V_1 and I_1''' produces the same value of information as the combination V_2 and I_1'. By choosing transaction arrangement V_2, the transactor reduces the cost of acquiring information of value OQ by $I_1'''-I_1'$. The saving in cost becomes available for consumption and exchange. The production of unchanged value of information with a smaller input of resources is a measure of the marginal productivity of transaction arrangement V_2.

An alternative measure of the productivity gain is the increase in the value of information obtainable with unchanged investment. The distance PR on the $\epsilon(V_2)$ axis is the increased value of information (in units of V_2) obtained with investment I_1'' and the choice of transaction chain V_2 instead of V_1. By using V_2 the transactor is able to trade at the ratio given by budget line 4 rather than the lower exchange ratio given by constraint 3. Thus, careful choice of transaction arrangements increases the productivity of a transactor's resources and his utility.

The results obtained in our world of three commodities apply to an n-commodity world in which there are numerous alternative transaction arrangements and many transactors. Each transaction arrangement can be represented by a curve on the left panel of Figure 2 showing the value of information obtained with a given investment of resources in information. Our first postulate implies not only that the curves differ substantially but that the optimal transaction arrangement for an individual transactor—the highest curve—involves indirect exchange, i.e., the use of an intermediate asset that reduces his cost of acquiring information about market opportunities. Our second postulate implies that what is true for the individual is true for most members of the social group. The transaction arrangements that are most productive for the individual are also most productive for a dominant portion of the group.

The assets used to reduce the cost of acquiring information and transacting are money for the group or society. The social and private productivity of a medium of exchange results from the use of transaction chains involving assets with these properties. Where knowledge of market opportunities and the qualities of goods is neither costless to obtain nor uniformly distributed, the use of money as a medium of exchange reduces the resource cost of exchanging. In the following section, we consider the social productivity of money and some implications of the analysis for the type of assets chosen as mediums of exchange and used as money.

III. The Social Services of Money

For individuals, money is a substitute for investment in information and labor allocated to search. By using money, individuals reduce the amount of information they must acquire, process, and store, and they reduce the number of transactions in which they engage to exchange their initial endowments for optimal baskets of goods. The use of money increases the welfare of each money user by reducing uncertainty, the length of transaction chains, and the variance of price ratios and by increasing expected wealth and time available for leisure. Whatever other services create a demand for the assets that serve as mediums of exchange, their use as mediums of exchange increases demand. Individuals find it advantageous to allocate part of their wealth to money.

What is true for individuals is in this case true for society. The convergence of optimal transaction arrangements generates an aggregate demand for the assets used as mediums of exchange. The increased demand to hold inventories of these assets (money) is independent of the previous uses of the assets and, of course, increases the (relative) prices of the assets. The average amount held in inventories depends on the prices of the assets held, the prices of alternative assets and, thus, on the relation between net marginal productivity and marginal cost.

Once inventories of money are held, payments and receipts are no longer synchronized. Lack of synchronization, however, does not explain the use of holding of money anymore than the holding of money explains the lack of synchronization. Both are a result of the superior productivity of indirect methods of exchange, the smaller resource cost of acquiring information and transacting in a monetary economy.[19]

The use of money encourages the development of the market system by lowering the costs of acquiring information and transacting. With the expansion of the market, opportunities increase for professional middlemen and specialized traders to exploit the partial and incomplete distribution of information about particular commodities. Specialized traders substitute for a wider and more general distribution of information. The use of money also affects the intertemporal allocation of resources. Deferred payments, borrowing, credit and the payments system expand when a standardized asset with well-known properties becomes available. The reason is that transactors become more willing to enter into contracts calling for deferred payment.

The magnitude of the net social productivity of money is not constant but varies with the degree of uncertainty abou market conditions, including exchange ratios and the quality of goods. Accelerated technological changes or innovations that change the qualities of goods and increase the number or types of goods raise the productivity of money. Large fluctuations in economic activity also raise costs of acquiring information and the productivity of money. Our analysis implies that the demand for mediums of exchange is higher in periods of rapid change than in periods of gradual or relatively steady change. The longer the period of steady gradual change continues, the lower the productivity of money and the smaller the demand for assets that reduce costs o acquiring information by serving as mediums of exchange.

A stationary state or a world of steady growth are the limiting cases of economic theory. Tastes, technology, anticipations population, and types of product are either invariant or change in a known, fully anticipated way at a steady rate. The marginal cost of acquiring information falls and in the limit approaches zero Transaction chains no longer differ by the saving of costs of acquiring information and differ only by the costs of transfer The main condition leading to the selection of a small group of assets as money and therefore the main source of the distinction between money and nonmonetary assets, disappears in the stationary state or world of steady growth.[20]

---

[19] In a well-developed market economy, most of the net *marginal* productivity of money probably results from the saving of costs of transacting, while the *total* productivity of money depends on the reduction in costs of acquiring information and costs of exchanging. The difference between our analysis and the usual analysis of "transaction costs" (see Baumol, Tobin (1956)) is that the *total* productivity of money in the familiar Baumol and Tobin analyses is almost the same as the *marginal* productivity in our analysis. The reason is that Baumol-Tobin transactors face fixed payment schedules, whereas our transactors optimize over all transaction arrangements. The Baumol-Tobin analysis has been used recently by Samuelson (1969), Clower and Johnson to equate the productivity of money with the "shoe leather" saved by avoiding trips to the bank or the market place. Their analyses understate the total contribution of money to wealth by neglecting the improvement in the opportunity set shown in Figure 2 and discussed in the text.

[20] In the terms of our formal analysis, the effect i conveyed by the distribution $\pi(E|P)$. The coordinate of the vector P summarizes the representative individual's information about market opportunities. The greater the frequency of change in market conditions

Hyper-inflation and hyper-deflation provide examples of changes in the marginal cost and marginal productivity of holding and using money. In both, the increased frequency of change in market conditions increases uncertainty and the variance of exchange ratios and thus raises the marginal productivity of money. But the marginal cost of holding money increases in hyper-inflation as the rate of inflation increases. Transactors are induced to choose transaction chains that avoid the inflation tax, even if such chains use more resources for transactions.

The search for new transaction chains and the abandonment of old is commonly mistaken for a return to barter. The process does not restore either double-coincidence exchange or the random selection of mediums of exchange. Transactors concentrate their search on those transaction chains that offer at least the same expected gain in wealth as existing mediums of exchange. The assets that replace existing money may have higher marginal costs of acquiring information and transacting. If so, they must have lower marginal holding costs than the existing money.

Our analysis implies that continued and accelerated hyper-inflation eventually comes to an end. The economy gradually adopts a new money and a new set of transaction arrangements. The process of search and social convergence is not costless for the individual transactors or the society. Governments or private producers can reduce the social cost of introducing new mediums of exchange and the transition to the new transaction arrangements by introducing new assets with properties similar to the old and effectively controlling the quantity produced.

A large deflation also increases uncertainty and induces transactors to reconsider transaction chains. Unlike hyper-inflation, the search for new transaction chains in a deflation does not replace existing mediums of exchange but adds new ones. The relatively high yield on money puts a premium on the search for assets that are close substitutes for the existing mediums of exchange—have similar information and transfer costs—and lower yield. New types of money evolve to supplement the existing money and increase the available stock of real balances. Adjustment to severe deflation, therefore, is not concentrated solely on the price level and rate of price change as in growth theory. New supplies of real balances are produced by the search for alternative transaction chains.

Our analysis of the use of money implies that both inflation and deflation are stable in a certain technical sense. In inflation new monies are substituted for old; in deflation supplementary monies are introduced. But the evidence, from periods of hyper-inflation and severe deflation alike, reveals the sizeable resource costs associated with the stabilizing mechanism. Hyper-inflation or deflation reach relatively high levels before transactors begin to replace existing money or add new money. The high costs that individuals are willing to pay before beginning to search for new or supplementary arrangements suggests the size of the benefits received from the use of a dominant medium of exchange.

The size of the net social productivity of money also depends on the assets selected as mediums of exchange. Once the community uses some assets as money, the private and social benefit can be increased by substituting claims against commodities for commodity money. Individuals gain from the use of substitutes for commodity money if the reduction of costs of acquiring information and transacting more than compensates for the increased

---

the poorer his information. In the stationary state his information increases without offset and approaches full information. The variances and covariances of the π function decline. See fn. 11.

variability of exchange ratios. Society gains because the use of claims and fiat paper money reduces the resources used to make exchanges in three main ways. First, paper money permits society to develop a fractional reserve money system and to produce the same nominal stock of money at lower resource cost. Second, the use of claims encourages the development of privately produced money and with it the development of the payments system. The cost of acquiring information about the qualities of paper money, whether produced by government or by private producers, is lowest if the paper money starts as a claim against commodity money. When information about the paper money becomes widespread, paper money retains the property of general acceptability even if the claim against commodities is removed. Third, paper money frequently lowers the resource cost of transfer and exchange. This somewhat less than general proposition recognizes that both costs and benefits are affected and that the size of the net benefit from the use of paper money depends on the prevailing monetary arrangements.

Monetary history offers numerous examples of changes in monetary arrangements that reduced marginal costs of information or transfer for the assets used in optimal transaction chains. Coinage is one of the earliest, and bank credit cards one of the most recent steps extending the range and use of mediums of exchange by reducing these costs.[21] Suppose, however, paper money is not introduced by a central bank or government but emerges in response to the public's search for optimal transaction chains. Many different producers are induced to issue paper money as a claim against commodity money. The social benefit resulting from the use of lower cost money is partly offset by the higher cost individuals pay to acquire information. The legislation of 1844 in England and of the 1860's in the United States that reduced the number and types of notes in circulation by restricting the right to issue notes are examples of institutional changes that raised economic welfare by reducing costs of acquiring information. The requirement of par collection of checks under the Federal Reserve Act is another example.

Nothing in our analysis implies either that society converges to a single medium of exchange or that the productivity of money and the contribution of money to wealth is limited to a single monetary asset. Different types of assets—some privately, some publicly produced—appear in the transaction chains adopted by members of the group and in the transaction chains of a single transactor at different times. These differences in the choice of transaction chain reflect differences in marginal cost that depend on the type of transaction and the transactor's information. Even in highly developed economies with extensive monetary institutions transactors can use specialized information to develop transaction arrangements that lower transfer costs by avoiding the use of money. Moreover, sectors of an economy that develop specialized information about the properties of particular assets often find it useful to develop specialized mediums of exchange.[22]

---

[21] Credit cards centralize information about deposit users, reduce a seller's cost of acquiring information, encourage the separation of payments and purchases and thereby increase (relatively) the use of deposits as a medium of exchange, lowering the ratio of currency to deposits. Lowering the costs of acquiring information and transferring via deposits lowers the net marginal productivity of deposit inventories. Average inventories of deposits are reduced; velocity increases.

[22] The development of the Federal funds market i an example. Corporate mergers offer examples of the way in which the allocation of human wealth (skilled specialists) reduces cost by avoiding money as a medium of exchange. By offering the owners of the merged firm deferred equity claims of various kinds, the purchaser reduce or defer the tax liabilities of equity owners

Our analysis suggests an extension of Gresham's law—cheap money drives out dear at fixed exchange rates—to the case of multiple mediums of exchange with variable or floating exchange rates. With floating exchange rates, stable monies drive out variable monies. Consequently, government or private producers desirous of maintaining the circulation of government and privately produced monies have found it desirable to provide arrangements for exchange of one money for the other on demand.[23] More importantly, issuers of privately produced money expand the market for their product by maintaining a fixed exchange rate. By doing so, they lower the users' cost of information and increase the demand for their product. Businessmen, bankers, and government officials have used similar reasoning to press for fixed exchange rates between national monies.

Throughout our analysis we have identified money with the medium of exchange. This usage has been criticized on two grounds. Tobin (1963) is one of the last in a long line emphasizing the existence of close substitutes for money. Tobin's criticism is part of a position discussed above, that ignores costs of acquiring information and attributes the total productivity of money to its role in synchronizing receipts and payments and reducing brokerage costs. Friedman and Schwartz (1970) also criticize the notion that money is a medium of exchange. They describe money as a "temporary abode of purchasing power" and argue that the term medium of exchange is an a priori notion devoid of empirical content. As an example of the deficiency of the medium of exchange concept, they cite the difficulty of using units of currency with large denominations or using checkable deposits in unfamiliar environments. Their argument is insufficient for the conclusion they have drawn. The use of checks and bills of large dedenomination often involves substantial costs of acquiring information and exchanging. On our analysis, it is not surprising that neither circulates in unfamiliar environments.[24]

Defining money as a temporary abode of purchasing power does not distinguish between properties of assets or between a monetary and a barter economy in a manner independent of the medium of exchange function. As our earlier discussion shows, transactors hold intermediate goods temporarily in a barter economy—as one of the items in a transaction chain—if their best information suggests that by doing so they can make more advantageous exchanges.

The recognition of the central role of a medium of exchange does not imply that the collection of assets that serve as medium of exchange is most appropriate for explaining movements of the general price level. A definition embracing a larger collection of assets is appropriate if there are close substitutes for the medium of exchange on the supply side. In this case, slight changes in relative prices reallocate output between the medium of exchange and other assets, so the collection of assets most useful for explaining changes in the general price level differs from the assets that serve as medium of exchange. However, even if evidence suggests that a broader collection is justified empirically and the term money is used to refer to the broader collection, the significance of the medium of exchange function and its im-

---

raising the owners' wealth and reducing the cost of acquisition.

[23] This is the rationale for the instant repurchase clause discussed by Pesek and Saving.

[24] Five and ten thousand dollar notes never circulated widely but were used primarily for transactions between banks where the marginal cost of acquiring information about the notes was low. With the development and extension of the Federal funds market and other lower cost means of making interbank transfers, the use of bills of large denomination declined markedly.

portance for explaining the productivity of monetary assets remains.

IV. Conclusion

The use of money remained puzzling as long as the theory of exchange was restricted to the case of perfect certainty, a world in which information about market prices and the qualities of goods and services is obtained at zero cost. Standard price theory eliminated the main reasons for the existence and use of money by confining choice to three options—production, consumption and exchange—and setting costs of acquiring information about exchange opportunities and qualities of goods to zero. With costs of executing transactions zero and information a free good, there are no costs of shopping to assure that exchanges take place at the most favorable prices and no benefits from reducing the resource cost of executing transactions and eliminating cross-hauling of commodities. Any asset is just as usable as any other for executing transactions and discharging obligations. As a result, attempts to explain the use of money generally accept some consequences of the use of money—such as lack of synchronization—as an explanation of the existence of money.

Our analysis extends the theory of exchange to include the cost of acquiring information about market arrangements, relative prices, or exchange ratios. Individuals search for those sequences of transactions, called transaction chains, that minimize the cost of acquiring information and transacting. The use of assets with peculiar technical properties and low marginal cost of acquiring information reduces these costs. Money is such an asset, and the private and social productivity of money are a direct consequence of the saving in resources that the use of money permits and of the extension of the market system that occurs because of the reduction in the cost of making exchanges.

Money is a substitute for the specialized market skills that are part of a transactor stock of knowledge or "human wealth." Resources allocated to search and to maintaining market information can be reallocated once money is used as a medium of exchange. Trade and the market system expand, and the economy becomes increasingly monetized. More and better quality information becomes available with the expansion of the market and the opportunities for division of labor that lead to the development of professional transactors such as brokers and specialists. The use of a unit of account further reduces the cost of making exchanges.

We do not attempt to explain the holding of money except in the trivial sense that the use of an asset implies that the asset is held. Nevertheless the analysis helps to explain why money continues to be used even in periods of accelerating inflation when the cost of holding money reaches very high levels.

By analyzing the productivity of money and relating the productivity of money to the analysis of trade in an exchange economy, we clarify the meaning of the phrase "the services of money" and suggest by implication the benefits that would accrue to the world economy from the use of a medium of exchange. Since the saving in brokerage costs or "trips to the bank," that is generally presented as the total product of money, is only the marginal product of money in our analysis, the gains from using money in international as in domestic exchange are considerably larger than is generally claimed. (See Baumol, Clower, Johnson, Samuelson (1969), Tobin (1965).) Once we leave the world of certainty and costless information, both the private and social productivity of money rise; the use of money ceases to be puzzling and becomes, instead, an implication of optimizing behavior.

REFERENCES

W. J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Quart. J. Econ.*, Nov. 1952, *66*, 545–56.

K. Brunner and A. H. Meltzer, "Further Investigations of Demand and Supply Functions for Money," *J. Finance*, May 1964, *19*, 240–83.

D. Cass and M. Yaari, "A Reexamination of the Pure Consumption Loans Model," *J. Polit. Econ.*, Aug. 1966, *74*, 353–67.

R. Clower, "Is There An Optimal Money Supply?," *J. Finance*, May 1970, *25*, 425–33.

M. Friedman, "The Quantity Theory of Money, A Restatement," in M. Friedman, ed. *Studies in the Quantity Theory of Money*, Chicago 1956.

———, "The Optimum Quantity of Money," in M. Friedman, ed., *The Optimum Quantity of Money, and Other Essays*, Chicago 1969.

——— and A. J. Schwartz, *The Monetary Statistics of the United States*, Nat. Bur. Econ. Res., *Stud. in Bus. Cycles, 20*, New York 1970.

J. C. Gilbert, "The Demand for Money: The Development of an Economic Concept," *J. Polit. Econ.*, Apr. 1953, *61*, 144–59.

H. G. Johnson, "Is There an Optimal Money Supply?," *J. Finance*, May 1970, *25*, 435–42.

J. M. Keynes, *The General Theory of Employment, Interest, and Money*, London 1936.

A. W. Marget, *The Theory of Prices*, New York I, 1938, II, 1942.

W. E. Mason, *Clarification of the Monetary Standard*, Univ. Park, Pa. 1963.

J. Niehans, "Money in a Static Theory of Optimal Payment Arrangements," *J. Money, Credit, Banking*, Nov. 1969, *1*, 706–26.

D. Patinkin, *Money, Interest and Prices*, 2d ed., New York 1965.

B. P. Pesek and T. R. Saving, *Money, Wealth and Economic Theory*, New York 1967.

P. A. Samuelson, *The Foundations of Economic Analysis*, Cambridge, Mass. 1947.

———, "Non-optimality of Money Holding Under Laissez Faire," *Can. J. Econ.*, May 1969, *2*, 303–08.

J. Schumpeter, *Business Cycles*, New York 1939.

M. Sidrauski, "Rational Choice and Patterns of Growth in a Monetary Economy," *Amer. Econ. Rev. Proc.*, May 1967, *57*, 531–44.

H. A. Simon, "Theories of Decision Making in Economics and Behavioral Science," *Amer. Econ. Rev.*, June 1959, *49*, 253–83.

G. Stigler, "The Economics of Information," *J. Polit. Econ.*, June 1961, *69*, 213–25.

H. Thornton, *An Enquiry into the Nature of the Paper Credit of Great Britain*, London 1802, reprinted New York 1965.

J. Tobin, "The Interest Elasticity of Transactions Demand for Cash," *Rev. Econ. Statist.*, Aug. 1956, *38*, 241–47.

———, "Commercial Banks as Creators of Money," in D. Carson, ed., *Banking and Monetary Studies*, Homewood 1963.

———, "Money and Economic Growth," *Econometrica*, Oct. 1965, *33*, 671–84.

The Supply of Rental Housing

By Frank de Leeuw and Nkanta F. Ekanem*

The elasticity of the supply of housing services is a critical unknown in our understanding of the housing market. Proposals to substitute housing allowances for some of the present housing subsidy programs cannot be evaluated unless we know the extent to which such demand subsidies would encounter an inelastic supply and serve to drive up rents.[1] The effects of programs such as public housing, which subsidize both demand and supply, also depend in part on the elasticity of supply in the private market.

The present paper is an attempt to use information on rent differences among metropolitan areas to estimate the elasticity of supply of rental housing. The central body of data is the U.S. Bureau of Labor Statistics (*BLS*) 1967 survey of the rental cost of housing units of fixed specifications in thirty-nine metropolitan areas. The *BLS* survey showed (confirming earlier surveys) sizeable differences in rent levels for comparable units in different parts of the United States. The paper relates these rent differences to regional variations in demand influences, such as number of households and average income, and supply influences, such as the costs of capital inputs and operating inputs. By utilizing outside information on demand elasticities, it then draws some inferences about the supply of housing services.

Studying differences among cities amounts to studying how housing markets behave in the long run, in the sense of having had ample time to adjust to basic market forces. The reason is that differences among cities in size, costs, tax rates, real income, and so on tend to persist for years or even decades. What we observe when we look at rent levels in different cities, therefore, is dominated by the cumulated effects of these long-term differences. The short-term behavior of the supply of housing is also of great interest, but estimating short-term responses requires different bodies of data from the ones used here.

The first part of the paper sets out a general supply-demand framework for analyzing the rent data. An important part of the framework is the distinction between several possible sources of rising supply price for housing services: rising prices of capital and operating inputs, diseconomies of scale in producing new stock, diseconomies of scale in maintaining existing stock, and diseconomies of scale in producing services from a given stock. The second section describes the data and the results of the statistical analysis. The final section draws some inferences about the supply of housing services and its implications for policies to improve housing conditions.

I. The Basic Framework

To understand the behavior of rent levels between cities, we begin with a demand-supply model of the housing market. A central concept in the model is the quantity of housing services. By quantity of housing services is meant a weighted average of space, heat, level of maintenance, nearness to work, and all other things of value provided by the housing stock. While it is not possible to measure the quantity of housing services directly, it is possible to measure the rental price of units which are identical in many of the characteristics determining the housing services they provide. The *BLS* survey provides such measures. In the model, the price of housing refers to such measures, and the quantity of housing refers to housing expenditures deflated by such measures.

---

* The authors are on the staff of The Urban Institute, Washington, D.C. The research and studies forming the basis for this publication were conducted pursuant to a contract with the U. S. Department of Housing and Urban Development. The views expressed are those of the authors and do not necessarily represent the views of The Urban Institute or the Department of Housing and Urban Development.

[1] See The President's Committee on Urban Housing, pp. 71–72, and Edgar Olsen, pp. 619–21.

On the demand side there is ample evidence that the quantity of housing services purchased by households in an area depends upon real income and upon the price of housing relative to prices of other goods and services.[2] Our analysis of expenditure and price data for 1960[3] confirms the importance of real income and relative price, and suggests an income elasticity for renter families of 1.0 or a little below, and a price elasticity somewhere in the neighborhood of -1.0.

On the supply side, there is much less to go on in the way of past work. In the short run, a relation between rents and vacancy rates has been discussed and observed,[4] and we have attempted to measure and test the role of vacancy rates in accounting for rent differences between cities; but in the long run new construction and removals tend to limit the variation of vacancy rates.

We assume that in the long run the supply price of housing services bears a positive relation to marginal costs—that is, to the costs to the housing and real estate industry of adding to the supply of housing services. Following this assumption, we would expect rents to be high in cities where land prices, wage rates of construction workers, financing costs, utility prices, wage rates of maintenance labor, or property taxes are high. We have constructed measures of these input prices in different cities, and grouped them into the two variables "average price of capital inputs" and "average price of operating inputs."

Marginal costs may also depend on the amount of housing services provided in an area. Our central interest is in this possible relation of industry costs to the amount of housing services. If housing services are characterized by a rising supply price, then we should expect higher real incomes (other things equal) to lead to higher rents per unit of service. On the other hand, if housing services are not characterized by a rising supply price, then higher incomes should not drive up rents per unit of housing service in the long run. As average income rises, the households in an area should be able to purchase more housing services at the same cost per unit of service.

There are several distinct ways in which a rising supply price for housing services could arise. One of them is inelastic supplies of inputs into the production of housing services. Land is often singled out in this regard; but there are other possibilities as well. Construction labor, maintenance men, security guards, painters, and exterminators—to name a few examples—are specialized inputs in the production of housing services which may command a higher price as the demand for them increases.

The other possible sources of rising supply price all come under the heading of diseconomies of scale. In distinguishing among the varieties of scale considerations involved in housing, it is helpful to factor housing services into the amount of housing services per unit of capital stock times the size of the capital stock, and to separate the capital stock into new stock and existing stock. Diseconomies of scale could arise in the production of services from capital stock and operating inputs, they could arise in the production of new capital stock, or they could arise in the maintenance of existing capital stock. An example of scale effects in the production of services is the additional output in relation to the additional resources associated with obtaining repairs promptly rather than after a long or unpredictable delay. An example of scale effects in maintenance is the additional output in relation to the additional resources associated with overhauling the plumbing, heating, or electrical systems of an old unit in order to keep it from falling below the standards demanded by most middle-income households.

---

[2] See Sherman Maisel and Louis Winnick; Richard Muth (1960); and Margaret Reid.

[3] Described in de Leeuw.

[4] See David Blank, Chester Rapkin, and Louis Winnick.

Muth's discussion of the supply of housing (1968) focuses primarily on the production of new housing units and the prices of inputs into new housing. The evidence which leads him to conclude that the long-run supply is highly elastic is consequently incomplete. It does not cover prices of operating inputs, the production of housing services from capital stock and operating inputs, or the maintenance of existing stock.

The present study is mainly concerned with the possible existence of an inelastic supply, or rising supply price, for housing services, not with its precise source. Nevertheless, available data permit some differentiation among possible sources of inelastic supply. If the supply of housing services is inelastic primarily because of inelastic supplies of inputs, then rent differences among metropolitan areas (for comparable units) ought to be related to demand factors only through the influence of input prices. That is, if there are no diseconomies of scale, rents should be significantly related to incomes and price levels when input prices are not included in the analysis; but rents should not be significantly related to demand factors if input prices are included.[5] Consequently, it will be of interest to test the relation of rents to demand variables both including and excluding variables measuring prices of inputs.

With respect to diseconomies of scale, all that available data permit is separate estimates of the relation of rents to services per household and of rents to number of households. The factoring of total services into services per household and number of households does not correspond very closely to the concepts discussed above of services per unit of capital stock and new and existing capital stock. However, it is likely that diseconomies of scale in the production of services would affect primarily the relation of rents to services per household, while diseconomies of scale in producing new stock or maintaining existing stock could affect either the relation

---

[5] In mathematical terms suppose we have the following reduced form relations. Equation (a) is derived from the equality of the supply and demand for housing services.

$$R = b_1C + b_2D + u \tag{a}$$

where R is rental price, C is an index of input prices, D is a demand influence such as income. Equation (b) is a reduced form relation derived from the equality of the demand and supply for inputs into the production of housing services.

$$C = b_3D + b_4Z + v \tag{b}$$

where Z is an influence on the supply of inputs, and u and v are disturbances uncorrelated with the independent variables in their respective equations. Let us assume there are no diseconomies of scale so that the supply schedule with respect to output price is perfectly elastic for given input prices. This implies that b_2 is equal to zero. The footnoted statement asserts that if b_2 is equal to zero but b_1 and b_3 are not equal to zero, then (i) a simple regression of R on D should yield a significant coefficient of D, and (ii) a multiple regression of R on C and D should yield an insignificant coefficient of D. The expected value of the ordinary least squares estimator of the coefficient of D is of course b_2 in case (ii), the multiple regression. In case (i), the expected value of the ordinary least squares estimator is

$$b_2 + b_1b_3 + b_1b_4\,m_{ZD}/m_{DD}$$

where m_{ZD} and m_{DD} are product-moments.

of rents to services per household or the relation of rents to number of households.

Two equations summarize the relationships of the present study: a demand equation for the average household and a supply equation in which it is convenient to express rental price as the dependent variable.

$$(1)\qquad S/H = g(Y, R/P)$$

$$(2)\qquad R = f(C, O, H, S/H, V)$$

where S/H represents the quantity of housing services per household, Y represents real income per household, R represents rent per unit of housing service, P represents the price of goods other than housing, C represents the price of capital inputs, O represents the price of operating inputs, H represents the number of households, and V represents the rental vacancy rate, an indicator of disequilibrium in a housing market.

The data available measure all of the variables in the two equations except services per household, S/H. It is therefore necessary to combine the two equations to eliminate S/H. We are left with a reduced-form relationship between rents per unit of service on the one hand and the following supply and demand influences on the other.

C = the price of capital inputs
O = the price of operating inputs
H = the number of households
Y = real income per household
P = the general price level in an area
V = the vacancy rate

For the first two variables, prices of capital and operating inputs, we expect a positive relationship with rents, while for the last, vacancies, we expect a negative relationship with rents. For the remaining variables—households, income, and the general price level—we expect a zero or positive relationship, with a positive relationship indicating a less-than-perfectly elastic supply.

The estimation of the reduced-form equation is itself of some interest, and has the advantage of avoiding the problems of simultaneous-equations bias. Furthermore, it is a relatively simple matter to combine the results of the reduced-form estimation with information from other studies on the parameters of the demand equation (equation (1)) and draw conclusions as to the behavior of the supply of housing services.

II. The Data and Results

Description and Evaluation of Data

The rental price measure and a number of other variables used in the analysis were taken from the *BLS* survey of the cost of "Three Standards of Living for an Urban Family of Four Persons" in the spring of 1967. The rental units priced in each metropolitan area for this survey met the following specifications:

> An unfurnished five-room unit (house or apartment) in sound condition; a completely private bath; a fully equipped kitchen; hot and cold running water; electricity, central or other standard heating; access to public transportation, schools, grocery stores and play space for children; and location in residential neighborhoods free from hazards or nuisances. [p. 42]

Units meeting these specifications had a wide distribution of rents within each area, and the *BLS* divided this distribution into thirds. The rent for the lowest of the three standards of living is an average of the lower third of the distribution of rents for these units. The rent for the moderate standard is the average of the middle third of the distribution, and the rent for the higher standard is the average of the upper third of the distribution. For the higher standard, some units were included in the distribution which had more than one private bath and extras such as "a central switchboard, secretarial, swimming or special facilities." The simple average of the low rents for thirty-eight metropolitan areas[6] was $1,048 per year or $87 per month. The average moderate rent was $108 per month, and the average high rent was $156 per month. The statistical work below deals with each of the three rental levels separately.

The wide distribution of sample rents within most metropolitan areas implies that either price varies greatly among the neighborhoods and individual units within an area, or that the units in the *BLS* sample vary greatly in the quantity of housing services they provide. The former explanation—price variation within rental areas—could arise if neighborhoods are imperfect substitutes for one another or if lags and imperfections in the rental market lead to large discrepancies between the actual and the equilibrium rent for a unit. The latter explanation—quantity variations among sample units—could arise if the *BLS* specifications of number of rooms, baths, access to schools, and so on, are not sufficiently detailed to define a narrow range of housing services.

A critical question in evaluating the results of the present paper is whether there are important errors of measurement in the *BLS* rental data considered as a measure of price. The discussion in the previous paragraph suggests that for individual units, the *BLS* measure might be in part a measure of quantity rather than a measure of price. If it is true that rents for sample units measure quantity as well as price, *and* if quantity variations do not cancel out in striking averages for metropolitan areas, then there are systematic errors in the rental price variable.

Errors of this kind, if they exist, seem likely to be correlated with real income per household in an area. The reason is that the units in the *BLS* sample are not controlled for "neighborhood quality"—crowding, quality of schools, local public services, and so on—and that variations among areas in these factors are probably related to variations in real income. Needless to add, these factors are important influences on the rent a household is willing to pay for its dwelling unit. It is possible, then, that a positive partial relation between the rental measure used in this study and real income is due to errors of measurement rather than to a less-than-perfectly elastic supply of housing services.

It seems much less likely that measurement errors in the rental price measure are correlated with the general price level than that they are correlated with real income. Crowding, quality of schools, and other elements of neighborhood quality may well vary with real income, but there is little reason to expect them to vary with the general price level. A positive partial relation between the rental measure used in this study and the price level therefore is probably a reliable indicator of a less-than-perfectly elastic supply.

A number of other variables were also drawn from the *BLS* study. The price of capital inputs was represented by the annual payments (not including insurance, taxes, or other operating costs) on a typical 1960–61, 25 percent-down mortgage for a "standard home" defined as a "five- or six-room house with one or one and one-half baths, a fully equipped kitchen, hot and cold running water, electricity, central or other heating," and neighborhood specifications the same as for rental units. This figure reflects differences in construction costs, financing charges, and land prices among metropolitan areas.

The prices of operating inputs were also taken from the *BLS* study. Operating costs included prices of fuel and other utilities, property taxes, insurance, repair and replacement costs, and certain other items as specified and priced for the "moderate" homeowner budget. Because there is some question as to the incidence of property taxes, two operating cost variables were tested, one including all of the property tax and the other including only half. The results below are for the second of these measures; but results using the first measure were very similar and would lead to the same conclusions as the second measure.

---

[6] The *BLS* study covered thirty-nine metropolitan areas, but we have not used data for New York because of the presence of rent controls.

The general price level in each metropolitan area is represented by the total cost of the moderate standard for renter families in the *BLS* study, but excluding the rental component of that cost. The rental component was excluded in order to avoid biased statistical results which might arise from having an imperfectly measured rental cost as both a dependent variable and as an important part of one of the explanatory variables.

The remaining variables were drawn from a variety of sources. Personal income estimates by metropolitan area are made by the U.S. Office of Business Economics. Rental vacancy rates were calculated by interpolating estimates in the periodic *Housing Market Analyses* of the Department of Housing and Urban Development, Federal Housing Administration. These were not available for all of the thirty-eight sample cities; for the missing areas, there was no way of estimating vacancies. Finally, estimates of the number of households in each metropolitan area were taken from *Sales Management.*

Statistical Results

In the discussion of the basic framework, it was argued that the existence of diseconomies of scale in producing housing services can be tested by comparing the relation of rents to demand factors (income and the general price level) excluding input price variables with the relation including input price variables. The following equations indicate that if we exclude input price variables, there is a clear positive relationship of rents to both income and the price level.

$$(3)\quad R_1 = \underset{(2.2)}{2.58} + \underset{(4.3)}{1.80P} + \underset{(3.7)}{.54Y} \quad (R^2 = .52)$$

$$(4)\quad R_2 = \underset{(2.3)}{2.85} + \underset{(3.8)}{1.72P} + \underset{(3.4)}{.53Y} \quad (R^2 = .46)$$

$$(5)\quad R_3 = \underset{(.7)}{1.15} + \underset{(2.3)}{1.43P} + \underset{(3.6)}{.78Y} \quad (R^2 = .37)$$

All variables are in logarithmic form; R_1, R_2, R_3 refer to the low, moderate, and high rent levels in the *BLS* study; P refers to the general price level excluding the rental component; and Y refers to median income per household deflated by the general price level. The numbers in parentheses below each coefficient are t-ratios; and the coefficient of determination for each equation is adjusted for degrees of freedom.

When we add to these relationships measures of input prices, the coefficients of the general price level fall but the coefficients of real income remain nearly unchanged. The equations follow.

$$(6)\quad R_1 = \underset{(2.3)}{2.72} + \underset{(2.9)}{1.13P} + \underset{(3.7)}{.55Y} + \underset{(3.3)}{.28C} + \underset{(1.8)}{.16O} - \underset{(-2.7)}{.04H} \quad (R^2 = .71)$$

$$(7)\quad R_2 = \underset{(1.8)}{2.50} + \underset{(2.6)}{1.26P} + \underset{(3.3)}{.60Y} + \underset{(2.4)}{.24C} + \underset{(.7)}{.07O} - \underset{(-2.4)}{.04H} \quad (R^2 = .60)$$

$$(8)\quad R_3 = \underset{(.8)}{1.78} + \underset{(1.6)}{1.24P} + \underset{(2.4)}{.70Y} + \underset{(.8)}{.14C} - \underset{(-.5)}{.09O} + \underset{(.2)}{.004H} \quad (R^2 = .34)$$

In this equation R_1, R_2, R_3, P, and Y refer to rents, prices, and real incomes as before; C refers to the price of capital inputs, O refers to the price of operating inputs, and H refers to the number of households. As before, all variables are in logarithmic form.

Even when input price variables are included, real income and the price of nonhousing goods have significant positive relationships to rent levels. The equations thus suggest that the supply of housing services per household is less than perfectly elastic, and that one of the possible varieties of diseconomies of scale is a source of the inelasticity. The relation to capital costs and operating costs is positive as expected, except for the insignificant relation of R_3 to operating costs. The relation of rents to the number of households is unexpectedly negative (but very close to zero) in two of the three equations, which suggests (following the discussion of the basic framework) that diseconomies in the production of housing services may be more important than diseconomies in the production of capital stock. Refitting the equations without the household variable gives these results:

$$(9)\quad R_1 = \underset{(4.1)}{4.39} + \underset{(2.1)}{.89P} + \underset{(2.4)}{.31Y} + \underset{(3.3)}{.31C} + \underset{(2.2)}{.20O} \qquad (R^2 = .66)$$

$$(10)\quad R_2 = \underset{(3.3)}{4.32} + \underset{(2.0)}{1.00P} + \underset{(2.2)}{.35Y} + \underset{(2.4)}{.27C} + \underset{(1.1)}{.12O} \qquad (R^2 = .54)$$

$$(11)\quad R_3 = \underset{(1.8)}{1.59} + \underset{(1.7)}{1.26P} + \underset{(3.1)}{.73Y} + \underset{(.8)}{.14C} - \underset{(-.6)}{.10O} \qquad (R^2 = .36)$$

They are similar to results including the household variable. The coefficients of real income are somewhat lower when the household variable is omitted, but the income and price variables continue to be statistically significant.

For thirty-five of the thirty-eight metropolitan areas it was possible to construct a crude measure of the rental vacancy rate as of the spring of 1967. Regressions including the vacancy rate did not change any of the results above. The simple correlation of the vacancy rate with rent levels is negative (ranging from $-.44$ for the low rent standard, to $-.20$ for the high standard) but the coefficient after taking into account income, price, and cost variables is practically zero in every case and never significant statistically.

One possible reason for the insignificance of the rental vacancy variable is measurement errors. The vacancy rate estimates are interpolations between *Housing Market Survey* benchmarks which are themselves estimates derived largely from postal vacancy surveys. Another possibility is that much of the variation in vacancy rates among metropolitan areas reflects differences in "normal" vacancy rates rather than different degrees of housing market tightness. It might be, for example, that in areas of rapid growth there is much more building ahead of demand than in areas of slow growth, and that normal vacancy rates tend therefore to be higher in rapid growth areas. Differences in normal turnover among housing types—single-family houses, apartments, boarding houses, and the like—might also cause differences among area-wide vacancy rates unrelated to housing market tightness. Whatever the explanation, the vacancy variable does not affect the principal finding of this section; namely, the significant influence of real income per household and nonhousing prices on rental price.

The data analyzed in this section thus suggest that the long-run supply of housing services is less than perfectly elastic. The next section derives more precise estimates of the long-run supply elasticity. The cause of the rising supply price does not appear to lie wholly in inelastic supplies of land or or other inputs, since the regression results indicate that upward shifts in demand factors (income and nonhousing prices) affect rents even after allowing for changes in input prices. The cause of the rising supply price appears to lie at least in part in one of the the possible sources of diseconomies of scale in the production of housing services.

III. Implications of the Results

The estimated equations show the influence on rents of a combination of supply factors and demand factors. From the equations alone it is not possible to disentangle the supply responses from the demand responses. By adding additional, outside information about the demand for housing, however, it is possible to derive supply relationships from the equations above. This final section of the paper presents this derivation and comments on the implications of the study for housing policies.

The Supply of Housing Services in the Long Run

To derive supply elasticities from the results of the previous section, we use information drawn from a review of cross section evidence on the demand for housing.[7] The review suggests an income elasticity for rental housing of about 1.0 or slightly below, and a price elasticity in the neighborhood of −1.0.

To combine this information with the equations above, we assume that the demand for rental housing services per household can be represented by the following relationship:

$$(12)\quad (S - H) = \alpha_1 + \beta_1 Y - \beta_2(R - P)$$

All variables are measured in logarithmic form; S represents the quantity of housing services; H represents the number of households; Y represents real income; R represents rents per unit of housing service; and P represents the price level of nonhousing goods and services. Because of the logarithmic form, $(S-H)$ represents housing services per household and $(R-P)$ represents rents relative to other prices. The demand study just referred to suggested that β_1 and β_2 are each approximately 1.0.[8] On the supply side we assume the following relationship:

$$(13)\quad R = \alpha_2 + \beta_3 C + \beta_4 O + \beta_5(S - H) + \beta_6 H$$

All variables are again measured in logarithmic form; R represents rents per unit of service; C represents the price of capital inputs; O represents the price of operating inputs; $(S-H)$ represents rental housing services per household; and H represents the number of households.

Setting β_1 and β_2 equal to 1.0 and solving this pair of equations for the rent variable leads to the following relationship:

$$(14)\quad R = \left(\frac{\alpha_2 + \alpha_1\beta_5}{1 + \beta_5}\right) + \left(\frac{\beta_3}{1 + \beta_5}\right) C + \left(\frac{\beta_4}{1 + \beta_5}\right) O + \left(\frac{\beta_5}{1 + \beta_5}\right)(Y + P) + \left(\frac{\beta_6}{1 + \beta_5}\right) H$$

This equation corresponds to the kind of relationship we have estimated empirically. It does not quite correspond to any of the relationships above, however, since it implies that the coefficient of real income and the coefficient of the general price level should be the same, whereas in the equations above they are permitted to differ. It is a simple matter to reestimate the equations above with this additional constraint, and the results of such estimates are as follows:

---

[7] See de Leeuw.

[8] In the demand studies, the price level variable did not exclude the price of rental housing, as it did in this study. But it is possible to rewrite the demand results in terms of the narrower price variable by approximating the general price level by a weighted geometric average of the housing price and the prices of other goods and services. If the income and price elasticities are equal in absolute magnitude, rewriting the demand results in this way does not change either of the elasticities.

$$(15)\quad R_1 = \underset{(2.0)}{2.33} + \underset{(3.9)}{.321C} + \underset{(2.1)}{.183O} + \underset{(4.0)}{.593(Y + P)} - \underset{(-2.6)}{.036H} \qquad (R^2 = .70)$$

$$(16)\quad R_2 = \underset{(1.5)}{2.06} + \underset{(2.9)}{.287C} + \underset{(1.0)}{.102O} + \underset{(3.6)}{.653(Y + P)} - \underset{(-2.4)}{.040H} \qquad (R^2 = .59)$$

$$(17)\quad R_3 = \underset{(.6)}{1.43} + \underset{(1.1)}{.178C} - \underset{(-.4)}{.066O} + \underset{(2.6)}{.745(Y + P)} + \underset{(.2)}{.004H} \qquad (R^2 = .35)$$

The coefficients of C in these equations correspond to $\beta_3/(1+\beta_5)$ in the general relationship above. The coefficients of O correspond to $\beta_4/(1+\beta_5)$, the coefficients of $Y+P$ correspond to $\beta_5/(1+\beta_5)$, and the coefficients of H correspond to $\beta_6/(1+\beta_5)$.

From these results we can derive relationships describing how landlords set their rent levels in the long run. The derivations imply the following landlord responses:

(a) the elasticity of rents with respect to capital costs is .8, .8, and .7 in the three equations;
(b) the elasticity of rents with respect to operating costs is .5, .3, and −.3 in the three equations;
(c) the elasticity of rents with respect to the amount of housing services per household is 1.5, 1.9, and 2.9 in the three equations; and
(d) the elasticity of rents with respect to the number of households is −.1, −.1, and zero in the three equations.

Another way of expressing these long-term relationships is to look at the supply of housing services instead of at the rent level as the dependent variable. To do this, we solve the basic supply equation for S the supply of rental housing services instead of R the rent level, and insert the empirical values of the β's. Looked at in this way, the empirical results imply that:

(a) the elasticity of the supply of housing services with respect to rent per unit of service is .7, .5, and .3 in the three equations;
(b) the elasticity of the supply of housing services with respect to the price of capital inputs is −.5, −.4, and −.2 in the three equations;
(c) the elasticity of the supply of housing services with respect to the price of operating inputs is −.3, −.2, and .1 in the three equations; and
(d) the elasticity of the supply of housing services with respect to the number of households is 1.1, 1.1, and 1.0 in the three equations.

For the most part these implications seem plausible. The negative coefficient of operating costs in the "high rent" equation is the one obvious exception. The supply elasticities with respect to rental price, which fall in the range .3 to .7, do not include any correction for the effects of possible measurement errors in the rental price measure. If there are errors of the sort discussed in the previous section, then these estimated elasticities with respect to rental price are too low.

The Aggregate Supply Elasticity and Housing Policies

Little can be concluded about policy questions of general interest from this single study. Nevertheless, it is useful to discuss the relation of the findings of this study to policies to increase the demand for housing by low-income households. Such policies include both earmarked assistance programs such as rent certificates, or programs of general income assistance which can be expected to be channeled in part into housing demand.

Superficially, the implication of this study would seem to be that programs to increase the demand for housing would in part serve to drive up rents. Some of the disbursements under such programs, that is, would be dissipated in higher rental prices rather than increasing the level of housing services. This section will conclude that the superficial implication is probably correct. But the conclusion rests on a number of additional assumptions and qualifications, which it is important to make explicit.

To begin with, it is worthwhile repeating that the results relate to long-run responses. Obviously, no conclusion is warranted from these results about the initial response of rents to an increase in demand.

Less obvious, but nevertheless important, is the point that the empirical results relate to the effects of changes in *average* demand per household on rents. The policies, in contrast, involve in part *shifts* in demand from one income-class to another. The implications of the empirical results for housing demand policy therefore involve the problem of how increases and decreases in demands in different subsectors of the housing market relate to the average responses studied in this paper.

By no means all of the demand assistance provided under a national income assistance or rent certificate program would simply consist of shifts in housing demand from one subsector of a metropolitan area rental market to another. A significant part of whatever demand changes were involved would consist of transfers out of high-income areas and into low-income areas. This is obvious when we consider the large differences between the average incomes of different metropolitan areas—within the thirty-nine *BLS* areas, for example, between Cedar Rapids ($4,218 per capita) or San Francisco ($4,401) at one extreme, and Durham ($2,728) or Orlando ($2,691) at the other.[9] Transfers from one area to another would presumably lead to higher rent increases in areas with a net inflow of income than in those with a net outflow. Furthermore, a significant part of any demand changes would be transfers from suburban homeowners to inner-city renters. They would involve, in other words, widely differing housing types, different geographic locations, and different political jurisdictions. It is hard to imagine a housing market flexible enough to accomodate such shifts without corresponding shifts in the relative price of housing services even over fairly long periods. Finally, an assistance program financed out of general revenue but earmarked in some degree for housing, such as a rent certificate program, would involve a net increase in housing demand as well as a shift in demand from one group to another.

Nevertheless, some portion of a program to increase low-income housing demand would consist of shifts in demand within metropolitan area rather than overall increases in area demands of the sort studied in this paper. The shift component would be relatively more important for a general income assistance program than for an earmarked program, and it would be more important for a locally financed program than for a national program involving transfers between metropolitan areas.

To relate the findings of the paper to shifts in housing demand, it is helpful to consider two extreme sets of assumptions about the submarkets of a metropolitan area housing market. At one extreme we can interpret what the empirical findings imply if the housing market of a metropolitan area is highly flexible. If there are very few barriers to conversion of housing units from one level of service to another, very few barriers to households moving within an area, and if cross-elasticities of demand between areas are high, then changing demand in any one sector of the housing market should have a very much smaller impact on overall rents than the average responses we have studied empirically. There is little reason in this case for expecting the price of housing services to be affected by redistribution of income within a metropolitan area.

---

[9] The numbers in parentheses are estimated 1967 per capita incomes in each metropolitan area (see Office of Business Economics).

At the other extreme, we can interpret what the empirical findings imply on the assumption that metropolitan area housing markets are highly inflexible—that is, that high costs, zoning laws, and other barriers prevent the conversion of housing units from one level of service to another, and that racial discrimination and other demand barriers restrict the movement of households from one neighborhood to another. In this case, the empirical findings about the response of average rents to income is in effect a sample of what goes on in the housing submarkets of a metropolitan area. It is a sample drawn mainly from the moderate-income sector of the market, and it uses average incomes in the entire market as a proxy to represent average incomes in that sector of the market. For the moderate-income portions of the rental housing market an increase of 1 percent in real income per household leads to an increase in rents per unit of housing service estimated to be in the neighborhood of one-half of 1 percent (see equations (6)–(11) above). If the low-income housing submarket is like the moderate-income sector in this respect, then this estimate of .5 should carry over into the low-income sector of the market. If demands and supplies are less price elastic in the low-income sector, a response greater than .5 might be expected; while if demands are less income elastic in the low-income sector, a smaller response might be expected.

The truth doubtless lies in between these two extreme sets of assumptions. On the one hand, racial discrimination, zoning laws, and costs of conversion of a housing unit from one level of service to another are important aspects of housing markets and make the "completely flexible" assumptions implausible. On the other hand, fairly rapid changes in the economic composition and level of housing services of particular neighborhoods have occurred in almost every big city. Unfortunately, our present understanding of housing market segmentation does not permit us to go beyond these broad generalizations.

Taken together, all of these observations about the rental housing market suggest that the response of rents to income changes estimated in this paper is probably an upper limit to the long-run change in rents which recipients of an income subsidy would face. The reduced-form response estimated in this paper is an elasticity in the neighborhood of .5 (see equations (6)–(11) above)—that is, a half a percent increase in rent for each 1 percent increase in income. Because of the possible statistical bias in this estimate and because of the possibility that flexibility among housing submarkets would reduce rent effects, the true effect is probably below .5. An elasticity range of .1 to .4 is probably as precise a judgment as can be made at present. This range would imply that a general income assistance program which raised average incomes in poor neighborhoods by 10 percent would raise rents in those neighborhoods by 1 to 4 percent in the long run.

The results of an income assistance program would of course depend on the provisions and the comprehensiveness of the program. In particular, a demand subsidy program specifically earmarked for housing could have severe rent effects. A program which raised average incomes in poor neighborhoods by 10 percent and channeled all of the increase into housing would raise recipients' housing demands by something like 40 percent.[10] An elasticity range of .1 to .4 would imply an average rent increase of 4 to 16 percent in this case. But since the overall increase in housing demand would be greater for an earmarked program than for a general income astance program, the range of .1 to .4 might be too low in this case. A significant fraction of disbursements under a housing allowance program might thus be dissipated in rent increases rather than leading to more housing services.

The contrast between a general income assistance program and an earmarked-housing subsidy leads to a final point. The issue of earmarked versus general income subsidies is usually discussed in terms of the preferences of recipients and taxpayers as to the allocation of total consumption. The point suggested by this study is quite distinct. It is that an earmarked program which causes a significant rise in the price of the earmarked service tends to be less effective in assisting total consumption levels than a program of general income assistance. A higher proportion of the disbursements are dissipated in price increases in the former case than in the latter. Consumption of the earmarked service by beneficiaries will of course rise more under an earmarked program than under a general program; but consumption of all goods and services by beneficiaries, after correction for price changes, will rise less.

The present study suggests that subsidizing the demand for low-income housing would drive up rents. There are undoubtedly other services for which a demand subsidy would be even more inflationary; medical services, for example, may be one of these. There are a great many goods and services, however, for which a shift in demand would probably encounter much less serious constraints on supply or separation of submarkets than in the case of housing. Hence, an income subsidy earmarked for housing is probably a less cost-effective way to raise real incomes of recipients than a general income subsidy.

REFERENCES

D. Blank, C. Rapkin, and L. Winnick, *Housing Market Analysis: A Study of Theory and Methods,* New York 1953.

F. de Leeuw, "The Demand for Housing: A Review of Cross-Section Evidence," *Rev. Econ. Statist.,* Feb. 1971, *53,* 1–10.

S. Maisel and L. Winnick, "Family Housing Expenditures: Elusive Laws and Intrusive Variances," *Proceedings of the Conference on Consumption and Savings,* Univ. Penn., 1959, *1,* 359–435.

R. Muth, "The Demand for Nonfarm Housing," in A Harberger, ed., *The Demand for Durable Goods,* Chicago 1960.

———, "Urban Residential Land and Housing Markets," in H. S. Perloff and L. Wingo, eds., *Issues in Urban Economics,* Baltimore 1968.

E. Olsen, "A Competitive Theory of the Housing Market," *Amer. Econ. Rev.,* Sept. 1969, *59,* 612–21.

M. Reid, *Housing and Income,* Chicago 1962.

"Sales Management Survey of Buying Power," *Sales Management,* June 1968.

The President's Committee on Urban Housing, *A Decent Home,* Washington 1968.

U.S. Bureau of Labor Statistics, *Three Standards of Living for an Urban Family of Four Persons,* Bull. 1570–5, Washington 1967.

U.S. Federal Housing Administration, Field Market Analysis Service, *Housing Market Analysis,* issued periodically for large urban housing market areas.

U.S. Office of Business Economics, "Metropolitan Area Income in 1967," *Sur. Curr. Bus.,* May 1969, 19–33.

---

[10] This estimate assumes that low-income families normally spend 25 percent of their incomes on housing.

The Composition of International Reserve Holdings: A Problem of Choice Involving Risk

By JOHN H. MAKIN*

This paper is directed to the gold-exchange world which existed before the advent of the two-tier gold market in March 1968. It considers the choice between dollars and gold as international reserves and views the economic forces underlying the substitution of dollars for gold as significant in determining the viability of a dollar-exchange standard.

Gold and dollars are both viewed as non-sterile assets with an expected return on gold and a certain return on dollars. In the case of dollars and gold, gold is often labelled as the "sterile" asset. This is incorrect since a substantial and variable forward dollar premium on gold futures has been paid in an active Zurich market, indicating a positive expected return on gold. This market quoted a price of gold six months forward, in terms of an annual premium on the market price, from 1961 to 1968. An individual willing to pay a premium in order to purchase "sterile" gold at about $35 an ounce six months in the future indicates that he attaches a positive probability to the chance of a rise in the dollar price of gold. Otherwise, why forego the opportunity to invest in interest-yielding assets? The average premium, from the second quarter of 1961 through the first quarter of 1968 was 6.39 percent.

The forward premium on gold as an index of the expected return on gold is a crucial factor, in the empirical analysis below, of the choice by central banks between dollars and gold. It proves to be a significant independent variable.

If we restrict consideration to short-term dollar liabilities (ninety days) then we have an asset (dollars) with a certain positive return (neglecting capital gains on ninety-day dollar assets) compared with an asset (gold) with an expected positive return.

The choice between dollars and gold is cast in terms of independent, utility-maximizing portfolio behavior. This is contrary to the view expressed by Fritz Machlup among others, that the holdings of reserves by officials are entirely politically determined:

> Official dollar holdings, on the other hand, are determined largely by political considerations. The central bank of a large industrial country does not look in the first place at the alternative costs and earnings of his asset mix, but rather on the advantages or necessities of international financial cooperation or non-cooperation. [p. 100]

A more recent view of this implied-interdependent-behavior view of reserve composition is expressed by Lawrence Officer and Thomas Willett. In this view, major responsibility for the operation (success or failure) of the current reserve currency system lies with a relatively small number

---

* University of Wisconsin-Milwaukee. This paper, constituting a large part of my doctoral dissertation, was completed with the financial support of the Federal Reserve Bank of Chicago. I am indebted to Arnold Zellner, Robert Mundell, Harry G. Johnson, Joseph Kvasnicka, and Tong Hun Lee and to two referees for helpful suggestions. Of course, any errors remain solely my responsibility.

of mutually identified major central banks. That operation is viewed as a "mixed-motive game." The reason the system has continued to function despite predictions by Peter Kenen and Robert Triffin of its likely failure, is that in the "crisis zone" (dollar liabilities exceed gold backing) when pressure arises to convert dollars into gold, "recognized interdependence" modifies behavior and cooperation is substituted for the independent alternative of a rush out of gold.

It is not my purpose here to suggest that the interdependent view of the system suggested by Machlup, Officer, and Willett is incorrect. My suggestion is that to view the international monetary system effectively as an n-party, x-sum game where the system survives as long as x is greater than zero is not very useful. It tells us only that after the system has collapsed that it had become a negative-sum game for all or a significant number of the participants and that they have, therefore, stopped playing.

A more useful view of the system lies somewhere in between the independence and the interdependence view. Since we can only *predict* under hypotheses derived from the independent view,[1] I suggest that we try to do so using the best possible specification of a model of utility maximizing behavior. Having made this attempt it is possible to infer the degree of recognized interdependence in the system from the success or failure of predictions based on hypotheses of independently determined actions. That is, we are saying that to the extent the actions of central banks are not independent they are interdependent.

From the success of predictions based upon the independent view of the system, we can make inferences about its durability *ex ante* rather than *ex post*. That is, if the predictive ability of utility-maximizating portfolio hypotheses improves over time, we may infer that the interdependence view of the system as held by its participants is deteriorating, and conversely.[2]

I. A Portfolio Behavior Model for Reserve Assets

We shall consider the choice between dollars, with a given rate of return, and gold, taken to bear an expected rate of return.[3] The model to be employed, based on the work of Donald Farrar, Haim Levy, and J. R. Hicks, is developed in detail in my May 1971 article. I shall only sketch the elements of the model here.

---

[1] Unless one can conceive of a determinate solution to an n-party, x-sum game where n is about ten and x is unknown *ex ante*.

[2] Other attempts have been made to analyze the sensitivity of holdings of reserves to rates of return. Specification problems of both the dependent and independent variables have plagued such work.

Phillip Bell regressed total dollar holdings relative to total international reserves on the covered and uncovered *U.S.-U.K.* interest differentials and found that the share of *U.S.* dollar liabilities in foreign exchange reserves was not sensitive to either factor. The problem is that both privately and officially held dollars relative to officially held reserves served as his dependent variable. Since two actors are involved in the behavior of the numerator (official and private) with only one (official) affecting the denominator, it is not legitimate to make inferences from the behavior of the ratio. Whose behavior is being tested? Also, as will be noted below, many countries don't hold most of their dollar assets as Treasury bills.

Helmut A. Hagemann has indicated (among many other conclusions) that "the interest rate on dollars" is not particularly significant in affecting foreign exchange reserves held by officials relative to their total reserves. He is attempting, however, to base inferences about the ratio of dollars to total reserves held by officials upon the behavior of the official foreign exchange/total reserve ratio. I have shown (1970) that this procedure may lead to the wrong inference about the direction of changes in the ratio of dollar assets to reserves from the change in the ratio of total foreign exchange to reserves. In the analysis below only the ratio of official dollar assets to official gold is used as a dependent variable. Discussion of the appropriate rates of return on dollars is included below as well.

[3] I employ as the index of concern about the gold value of the dollar, the price of gold futures for contracts denominated in dollars in the Zurich market, as taken from *Pick's Currency Yearbook*. The "price" is expressed as a premium over the present price of gold in annual percentage terms. A rise in the premium is taken as evidence of a rise in the expected return on gold that will be translated into pressure to get out of dollars and into gold. We shall assume that the expectations formed in the private market generalize to officials.

Taking utility, U, to be derived from the returns, R, from an asset where $U(R)$ is a cubic[4] of the form

$$U(R)=R+bR^2+cR^3,$$

expressions for the marginal expected utility (M_1 and M_2) from each of two assets in a two-asset portfolio are derived. These expressions are given by equations (1) and (2).

$$\begin{aligned}(1)\quad M_1 &= F_E R_1 + F_V(2\sigma_1^2 X_1 + 2\rho_{12}\sigma_1\sigma_2 X_2) \\ &\quad + F_Q(3\sigma_1^3 X_1^2)\end{aligned}$$

$$\begin{aligned}(2)\quad M_2 &= F_E R_2 + F_V(2\sigma_2^2 X_2 + 2\rho_{12}\sigma_1\sigma_2 X_1) \\ &\quad + F_Q(3\sigma_2^3 X_2^2)\end{aligned}$$

where $F_E>0$, $F_V \gtrless 0$,[5] and $F_Q>0$ are the partial derivatives of the expected utility from returns with respect to the expected rate of return, E, its variance, V, and skewness, Q, respectively. Also, R_i= the rate of return on asset $i(i=1, 2)$; σ_1^2= variance of return on asset i; X_i= number of dollars spent on asset i; ρ_{ij}= correlation coefficient between R_1 and R_2; and σ_1^3= third central moment about the mean of R_i (which rises with positive skewness.)

Expected utility from the portfolio is maximized when marginal expected utility from asset 1 (which we shall take as gold or items with the equivalent of a gold-value guarantee), equals marginal expected utility from asset 2 (dollars). Given a budget constraint of OK_j for country j, and taking the return on dollars to be fixed, ($\sigma_2=0$) the portfolio situation for any country "j" is represented in Figure 1, where now

$$(3)\qquad M_2 = F_E R_2$$

since R is taken to be fixed. As shown in Figure 1, M_1 remains as given by equation (1). Moving from left to right indicates that the share of asset 1 in the portfolio is rising. The reverse raises the share of asset 2. The equality of M_1 and M_2 gives a utility maximum at Z_1. We have drawn M_i convex[6] to the OK_j axis to indicate positive skewness for asset 1.

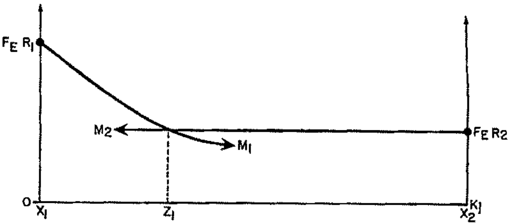


FIGURE 1

From Figure 1 we expect that holdings of asset 1 (gold) will rise with a rise in the return on gold and a rise in the positive skewness of the return on gold and fall with a rise in the variance of the return on gold for relatively low values of R_1 (see fn. 5). Holdings of asset 2 (dollars) will rise with R_2.

Letting G signify asset 1 (gold) and L signify asset 2 (dollar liabilities held by central banks of the United States), our portfolio behavior model implies the desired ratio of dollars to gold is given by:[7]

---

[4] This case is discussed by Giora Hanoch and Levy. See also Levy. In specifying a model which depends only upon maximization of the utility derived from returns from holding assets, we are focusing upon the question of reserve *composition*. This is separate from the question of the level of reserves which is effectively taken as exogenously given here. The problem of rates of return affecting the scale or level of reserves is dealt with by employing the *ratio* of dollars to gold as a dependent variable.

[5] In cubic utility function like the one employed here, the sign of F_V is negative for lower values of R and positive for higher values of R. That is, the attitude toward risk changes from risk averse to risk preference. The implications of this for the empirical results are noted below.

[6] If skewness were negative, M_1 would be bowed upward. Note that here we do not consider the possible case of an unstable equilibrium where a rising M_1 might cut M_2 from the bottom to the right of Z_1.

[7] Assuming the relationship is linear. Experimentation with a log-linear specification did not dramatically alter the results.

$$(4)\quad (L/G)^* = \alpha + \beta_1 R_G + \beta_2 VG + \beta_3 QG + \beta_4 R_L$$

where R_G = the return on gold; VG = the variance of the return on gold; QG = the skewness of the return on gold; and R_L = the return on dollars. We expect $\beta_1 < 0$, $\beta_2 > 0$, $\beta_3 < 0$, and $\beta_4 > 0$.

II. Testing the Model

A. Definition of Symbols Used in Presenting Results

a. Time Period: June 1961 through March 1968 (81 observations)

b. Dependent Variables:

D/G = time and demand deposit dollar assets of central banks held in *U.S.* banks/gold holdings
B/G = *U.S.* Treasury bills/gold
L/G = total dollar assets/gold (All are adjusted for SWAPS and Roosa Bonds where such facilities are used; indicated by D/GA, B/GA or L/GA)

c. Independent Variables:

R/G = return on gold (forward premium annualized)
R_{GAV} = average return on gold for six previous months
R_{GW} = a weighted average of the return on gold for six previous months (weights: .03, .10, .18, .24, .28, 17)[8]
VG = variance of R_{GAV}
QG = skewness[9] of R_{GAV}
VGW = variance of R_{GW}
QGW = skewness of R_{GW}
R_{ED} = return on Euro-dollars (average of daily figures)
R_{EDAV} = average return on Euro-dollars for six previous months
R_{EDW} = weighted average of return on Euro-dollars for six previous months
R_{US} = *U.S.* Treasury bill rate (average of daily figures)
ρEG = correlation coefficient between return on gold and Euro-dollars for six previous months
$L.D.$ = *U.S.* liquidity deficit
P^* = rate of change of prices in the United States minus the rate of change of prices in country being considered (where available; otherwise just the rate of change of prices in the United States = P)

d. Absolute *t*-ratios are in italics beneath the coefficients with which they are associated.

DW = Durbin-Watson statistic
$\bar{R}^2$ = percentage of explained variance adjusted for degrees of freedom

---

[8] These weights are chosen to give more weight to older observations. It is found that independent variables weighted in this manner serve as better predictors in cases where there is evidence of a slow process of portfolio adjustment. Over a distributed lag period six months in length, three weighting patterns were tried: giving all observed values equal weight; giving older observations more weight; and giving more recent observations more weight. (The six-month period was chosen in order to lose as few observations as possible. Experimenting with a period of twelve months did not improve results.) The results giving the highest $\bar{R}^2$ are reported in Table 1. In some cases, as reported in Table 1, single values of the independent variables gave better results than distributed lags. A formal method proposed by Shirley Almon is available for calculating a set of weights. It should be noted, however, that problems with the Almon procedure have been observed by G. S. Maddala. He found that a high degree of multicollinearity existed between the "synthetic" (weighted) independent variables arrived at by the Almon procedure and that weights estimated were sensitive to the length assumed for the lag and to the imposition of zero restrictions at -1 and $m+1$ where the lag runs $1 \ldots m$. These reservations about the Almon procedure together with the fact that results were not dramatically sensitive to alterations in the weighting pattern applied to past observed values of the independent variable, led to the use of the informal procedure described above.

[9] Note that here the index of skewness is the third central moment divided by $(s^2)^{3/2}$ to adjust for scale of measurement.

λ = partial adjustment coefficient

F = the F-statistic for test of the hypothesis that $\beta_1=\beta_2=\beta_K=0$

$$F = \frac{R^2/K}{(1-R^2)/(T-K-1)}$$

with $(K, T-K-1)$ degrees of freedom,

K = the number of independent variables;

T = the number observations;

$$R^2 = \frac{\text{regression sum of squares}}{\text{total sum of squares}}$$

In this study the Euro-dollar interest rate is taken to be a proxy for the return on dollar assets in a number of cases.[10] As a more sensitive index of confidence in the gold value of the dollar, we used the annualized forward premium on gold futures in the Zurich market as reported in *Pick's Currency Yearbook*. This is taken as the return on gold.[11]

An additional note regarding recorded gold and dollar holdings by foreign central banks is in order. Where Swap facilities drawn by the U.S. Federal Reserve are operative, they are subtracted from recorded dollar holdings and added to gold. Roosa Bonds are added to gold stocks. Both facilities are treated as a substitute for conversion of dollars into gold.[12]

B. The Reserve-Asset Ratio Adjustment Model

Since the data used are monthly, it was assumed following a partial adjustment hypothesis, that all desired adjustment of portfolios is not accomplished in a given time period. This gives:

$$(5) \quad \Delta(L/G) = \lambda((L/G)^*_t - (L/G)_{t-1})$$

where $(L/G)^*_t$ is the desired ratio of dollars to gold at time t.

In the empirical tests the initial specification of $(L/G)^*_t$ differs slightly from the expression given by equation (4) in including the correlation coefficient between the return on gold and the return on dollars, ρ_{EG}. In cases where the variance and skewness of the return on gold were calculated over a six-month period, since the return on dollars varied as well over the same period, it was possible to calculate the correlation coefficient between these two returns for each bloc of six months in the time period examined.[13] Therefore we specify:

---

[10] The reason for this is that a large part of dollar assets held by official foreigners are held in the form of time deposits at *U.S.* commercial banks. The rates paid on these deposits are not subject to Regulation Q. While the actual rates paid are not publicly available, I have been assured by an economist with a bank doing considerable business both in the market for foreign official deposits and for Euro-dollars that the rate moves closely with the Euro-dollar rate. For this reason the Euro-dollar rate is taken as a proxy for the rate paid on foreign official time deposits at *U.S.* commercial banks.

If central banks add to their claims on Euro-banks as well as on *U.S.* banks when R_{ED} rises, we must assume that *U.S.* banks keep their rates at a constant level of competitiveness with Euro-dollar rates in order to avoid the possibility of a wrong sign on $\hat{\beta}_4$. If such competitiveness is maintained $\hat{\beta}_4$ only understates the responsiveness of dollar denominated claims to an increase in dollar rates. This only strengthens the suggestion below that the sensitivity of dollar holdings to changes in dollar rates is above that for gold.

[11] Strictly speaking, the premium itself is an index of the expected return on gold. Here, we shall be taking a weighted average of six observations on the premium as an index of the expected return on gold. This is done in order to enable calculation of the variance and skewness of the moving weighted average of the premium on gold futures contracts expressed in dollars.

[12] Empirical evidence supporting this view is presented in Makin (1971). The exchange guarantee on Roosa Bonds makes them equivalent to gold if the monetary authority of a country pegs to gold (Switzerland). In the case where a monetary authority foresees a likelihood of raising the value of its currency in terms of both dollars and gold, Roosa Bonds are actually preferable to gold. For a country pegging to the dollar and contemplating a devaluation, Roosa Bonds are inferior to dollars and gold. Presumably, countries in the last category will not agree to hold such assets and will be more likely to insist on gold (France). I am indebted to a referee for these points.

[13] This raises the possibility that variance and skewness of the return on dollars will affect portfolio composition. Tests of this possibility did not yield statistically significant results.

$$(6)\quad (L/G)^{*}_{t} = \alpha + \beta_1 R_{Gt} + \beta_2 VG_t + \beta_3 QG_t + \beta_4 R_{Lt} + \beta_5 \rho_{EGt} + u_t$$

Then:

$$(7)\quad (L/G)_t = \lambda\alpha + \lambda\beta_1 R_{Gt} + \lambda\beta_2 VG_t + \lambda\beta_3 QG_t + \lambda\beta_4 R_{L_t} + \lambda\beta_5 \rho_{EGt} + (1-\lambda)(L/G)_{t-1} + \lambda u_t$$

where, given $0 \leq \lambda \leq 1$, we expect $\lambda\beta_1 < 0$, $\lambda\beta_2 > 0$, $\lambda\beta_3 < 0$, $\lambda\beta_4 > 0$, and $\lambda\beta_5 < 0$ if $(L/G) < 1$ and $\lambda\beta_5 > 0$ if $(L/G) > 1$.[14]

It can be shown that the presence of the lagged endogenous variable in an equation estimated by ordinary least squares always results in a Durbin-Watson statistic biased toward 2.00. (See Marc Nerlove and Kenneth Wallis.) Estimating equation (7) directly produced results in ten of the seventeen equations reported in Table 1 where the hypothesis of serial correlation could not be rejected at the .05 level of significance. In view of the bias noted by Nerlove and Wallis, the hypothesis of no correlation in the error terms in these cases surely must be doubted. The problem which serial correlation introduces is to bias the estimate of the coefficient attached to the lagged dependent variable. This has been shown by E. Malinvaud. In order to correct for this bias, following Malinvaud and Zvi Griliches, we assume that the time dependence in the error term of equation follows a first-order Markov process in cases where the Durbin-Watson statistic obtained from estimation of equation (7) is in or close to the region where a hypothesis of serial correlation cannot be rejected. That is, we write:

$$(8)\qquad v_t = \theta v_{t-1} + e_t$$

where $v_t = \lambda u_t$, and $e_t \sim N(0, \sigma^2)$. It is assumed that v_t is independent of all explanatory variables in equation (7) save for the lagged dependent variable. Substituting into equation (7) an expression for λu_t implied by equations (7) and (8), and rearranging terms we get:

$$(9)\quad (LG)_t - \theta(L/G)_{t-1} = \lambda\alpha(1-\theta) + \lambda\beta_1(R_{Gt} - \theta R_{G_{t-1}}) + \lambda\beta_2(VG_t - \theta VG_{t-1}) + \lambda\beta_3(QG_t - \theta QG_{t-1}) + \lambda\beta_4(R_{Lt} - \theta R_{L_{t-1}}) + \lambda\beta_5(\rho_{EGt} - \theta\rho_{EGt}) + (1-\lambda)(L/G_{t-1} + \theta L/G_{t-2}) + e_t$$

This procedure yields consistent estimates of $\lambda\beta_i$ $(i=1\ldots5)$ and of $(1-\lambda)$. Since a consistent estimate of λ can be obtained from $(1\widehat{-}\lambda)$, we can get a feel for relative speeds of adjustment by different central banks. A priori, we of course expect $0 \leq \lambda \leq 1$. As estimated λ falls we may infer a more costly adjustment process.

C. *Remarks on the Data and Estimation Procedure*

Before proceeding directly to the estimation of the model, it should be observed that L, dollar assets held by central banks, are not a strictly homogeneous entity. Some dollar assets held by central banks are in the form of *U.S.* Treasury bills. Some are in the form of time and/or demand deposits with private *U.S.* commercial banks and are not subject to Regulation Q. It is possible that in some cases only a subportion of dollar assets will be managed as part of a portifolio with gold. The costs of managing a particular type of dollar asset or subportion will be lower if such management is carried on with a single institution, either the U.S. Treasury or a *U.S.* commercial bank. Management of a subportion of dollar assets relative to gold is most likely to be the case for a large central bank with a high ratio of gold to dollars. The reason for this is that utility maximizing behavior could involve large enough shifts into gold to precipitate a run

---

[14] This follows from a consideration of M_1 and M_2 as specified by equations (1) and (2).

TABLE 1—ESTIMATED EQUATIONS[d]

Country	Dependent Variable	Constant $\lambda\alpha$	Expected Return on Gold R_{GW}	Variance VGW	Skewness QGW	Return on Dollars R_{EDW}	Liquidity Deficit or P^*	Correlation Coefficient or ρ_{EG}	Lagged Dependent Variable $(1-\lambda)$	$\bar{\lambda}$	$\bar{R}^2$	F	D-W
France:[a]	L/GA	+.2360 *4.74*	$-.0217R_G$ *4.12*			$-.0069R_{ED-1}$ *1.02*	−.000025 *2.69*		+.6916 *11.45*	+.3084	.95	391.04	2.02
Germany:	D/GA	+.0239 *2.59*	$-.00709R_{GW}$ *2.67*		$-.00489QGW$ *4.15*	$+.00558R_{EDW}$ *1.25*		−.00032 *.09*	+.0963 *1.06*	+.9036	.34	7.69	1.88
Germany:	D/GA	+.0235 *2.40*	$-.00710R_{GW}$ *2.66*	$+.00137VGW$ *.11*	$-.00488QGW$ *4.08*	$+.00565R_{EDW}$ *1.25*		−.00040 *.11*	+.0967 *1.06*	+.9033	.34	6.32	1.88
Italy:	D/GA		$-.01004R_{GW}$ *3.74*		$-.003946QGW$ *3.21*	$+.01543R_{EDW}$ *4.03*			+.7725 *13.41*	+.2275	.82	190.26	1.93
Italy:	D/GA	−.0026 *.23*	$-.01027R_{GW}$ *3.74*	$+.004158VGW$ *.34*	$-.00375QGW$ *2.91*	$+.01626R_{EDW}$ *3.44*		−.002238 *.61*	+.7797 *13.17*	+.2203	.82	55.02	1.93
Switzerland:[a]	D/GA	+.0014 *.19*	$-.005271R_{GAV}$ *2.34*	$+.00461VGAV$ *.66*	$+.00054QGAV$ *.36*	$+.0064R_{EDAV}$ *1.88*		−.00370 *1.53*	+.7501 *8.94*	+.2499	.78	44.09	1.94
Switzerland:[a]	D/GA	+.0030 *.43*	$-.00498R_{GAV}$ *2.47*			$+.0058R_{EDAV}$ *1.84*		−.00326 *1.40*	+.7563 *11.30*	+.2347	.78	67.20	1.91
Denmark:	D/G	−.0783 *1.41*	$-.02936R_{GW}$ *2.38*			$+.0663R_{EDW}$ *2.89*	−.000026 *1.08*		+.8651 *14.83*	+.1349	.91	185.21	1.94
Sweden:	L/G		$-.05273R_{GW}$ *2.02*	$+.0348VGW$ *.29*	$-.02937QGW$ *2.24*	$+.1112RR_{EDW}$ *2.30*		−.03665 *.95*	+.9290 *20.31*	+.0710	.89	3279.4	2.02
Spain:[a]	L/G	+.08256 *1.43*	$-.02069R_{GAV}$ *1.90*	$+.00062VGAV$ *.018*	$-.00631QGAV$ *.78*	$+.0179R_{EDW}$ *1.00*		−.02599 *1.87*	+.8553 *15.35*	+.1447	.88	88.72	2.00
Spain:	D/G	+.05328 *1.45*	$-.01091R_{GAV}$ *2.12*		$-.004223QGAV$ *1.06*	$+.00828R_{ED-1}$ *1.01*		−.01668 *2.34*	+.8649 *16.65*	+.1351	.91	142.05	1.94
Belgium:[a]	B/GA	−.00856 *1.47*	$-.003128R_{G-1}$ *2.40*			$+.01238R_{US-1}$ *3.97*			+.6234 *7.95*	+.3766	.80	101.99	1.88
Japan:	B/G	+.6047 *3.03*	$-.08688R_G$ *3.37*			$+.1171R_{US-1}$ *2.44*			+.7478 *11.67*	+.2522	.74	71.64	1.80
Netherlands:	D/GA	+.00808 *.72*	$+.002134R_G$ *1.31*			$-.00585R_{ED}$ *1.50*	$+.00074P^*$ *.29*		+.7383 *9.45*	+.2617	.56	24.25	1.86
Netherlands:[a]	B/GA	+.06513 *3.20*	$+.00412R_{GAV}$ *1.01*		$+.037R_{USW}$[b] *2.17*	$-.03829R_{EDAV}$ *2.07*			+.3586 *3.24*	+.6414	.26	6.62	1.92
Canada:	L/G	+.5572 *2.74*	$-.0633R_{GW}$ *2.44*			[c]			+.8093 *13.02*	+.1907	.83	190.57	1.85
United Kingdom:	D/GA		$-.01836R_{GW}$ *3.08*			$-.02293R_{EDW}$ *2.94*			+.9657 *20.66*	+.0343	.89	297.66	2.01
Austria:[a]	L/GA	+.1815 *2.24*	$-.05787R_{GW}$ *.81*			$-.01982R_{ED-1}$ *1.56*			+.8668 *15.36*	+.1332	.95	449.04	1.91

---

[a] The estimated equation is of the form given by equation (7) since the hypothesis that $\hat{\theta}=0$ cannot be rejected at the .01 level. All other estimated equations are of the form given by equation (9)

[b] R_{USW} = a weighted average of *U.S.* Treasury bill rates for the previous six months. See text for weights.

[c] Omitted as below F-level cutoff by program.

on *U.S.* gold stocks. If this forced a devaluation of the dollar in terms of gold, these (absolutely) large holders of dollars could lose on their dollar holdings as well as impair the effectiveness of a main international medium of exchange and store of value. The role of recognized interdependence comes into play.

The use of the return on Euro-dollars as an independent variable representing the return on dollar assets requires comment. Given the exemption from Regulation Q for deposits of foreign official institutions at *U.S.* banks, a bargaining process determines the behavior of dollar deposits (*ceteris paribus*) when the Euro-dollar rate rises. As was noted above, rates paid by *U.S.* banks do move with Euro-dollar rates making dollar assets more attractive relative to gold. But since the rates paid are based on outcome of a bargaining process we would expect that for countries with loose official ties or for countries openly resentful of *U.S.* dominance in the international monetary system, the process may fail. Then a rise in Euro-dollar rates would mean a move into Euro-dollars from *U.S.* dollars. However, this still helps the dollar exchange standard as dollar assets become more desirable relative to gold (*ceteris paribus*).

Bearing in mind these qualifications about the dependent and independent variables, equations were fitted with L/G, D/G, and B/G, adjusted for Swaps and Roosa Bonds where such instruments were in use, as dependent variables. These were regressed on independent variables, R_{GW}, VGW, QGW, R_{EDW}, and ρ_{EG} calculated by giving heavier weights to older observed values in arriving at expected returns, variance and skewness. (See fn. 8 for a discussion of the weighting procedure.) Each of the dependent variables was also regressed on simple averages of observed values of the independent variables as a check on the effect of altering the weighting pattern applied to the independent variables. For countries where neither variance nor skewness was in any case significant they were dropped and the equation rerun. In these cases a check was made to see how well the current value of the forward premium on gold futures in the Zurich market served as an index of the expected return on gold, along with the current value of the return on dollars, taken by itself, as the return on dollars. These independent variables were also tried, lagged one period.

Further, if either variance or skewness proved significant at the .05 or the .01 level both were included in one version of the estimated equation reported to check for the sign of the coefficient associated with each and for its possible explanatory power.

D. *Estimation Results*

The most striking feature of the results presented in Table 1 is the significance of the index of the expected return on gold in explaining the behavior of the ratio of dollars to gold and gold substitutes. It is highly significant[15] with the correct sign in nine of the thirteen cases.[16]

---

[15] Throughout this study it is important to note that the use of the t-ratio as a test of the level of significance means more precisely that we are testing the hypothesis, given that the estimated coefficient carries the correct sign, that it is significantly different from *zero*. This is done in order to specify as the null hypothesis the claim of those who assert that economic arguments have no impact on portfolio holdings, by central banks, of dollars relative to gold. We then test for coefficients which permit rejection of the null hypothesis. This procedure implies a testing strategy designed to minimize the chance of accepting a false hypothesis that a coefficient is different from zero when it actually is not. Of course in granting the critics of the economic approach the null hypothesis that $\partial(L/G)/\partial R_i=0$, $(i=G, ED)$ and minimizing the chance of accepting a false hypothesis, we run the risk of rejecting a true hypothesis that a coefficient actually has a true value closer to zero than, say, 1.96 times the standard error of the estimate of that coefficient. For estimated coefficients which carry the anticipated sign at, say, the .05 level of significance, a 95 percent confidence interval for that coefficient does not include zero.

[16] It should be noted that D (time and demand dollar

The impact of the return on dollars carries the correct sign for eight of the thirteen countries considered, although it is highly significant in only five of the cases. Some of the problem here may lie with the use of Euro-dollar rates as a proxy for the (unavailable) actual return on dollars held by central banks at *U.S.* commercial banks. In the cases of Belgium and Japan, both holding about 40 percent of their dollars as Treasury bills, the identifiable return on this form of dollar assets, the Treasury bill rate, performs well.

In the cases of France and Austria, a possible explanation for a fall in dollar assets held when the return on dollar assets (as indicated by the Euro-dollar rate) rose, particularly in the case of France in the 1965–67 period, is resentment felt against the role of the dollar as an international reserve asset. When Euro-dollar rates rise, it appears that these countries move their short term dollar assets directly into the Euro-dollar market. The reliability of this inference is low, however, since the coefficients attached to the return on dollars for these countries in Table 1 are not close to being significant at the .05 level. In the case of Austria it may simply be more convenient to manage its dollar assets with an eye to the Euro-dollar market instead of by negotiation with *U.S.* commercial banks.

The highly significant coefficient with correct sign attached to the expected return on gold in the case of France is further evidence of independent behavior. No such claim can be made for Austria.

It was thought that since Canada holds most of her short-term dollar assets as Treasury bills that regression of B/G o
R_{us} and R_G would produce results consistent with our theory. The t-ratios remaine
low, however, in fact below those obtaine
with L/G as the dependent variable. It i
not, however, particularly surprising t
find the independent, portfolio hypothesi
failing in the case of Canada. Close tie
with the United States would lead one t
expect more likelihood of interdependen
behavior.

The wrong signs associated with the return on dollars and the expected return o
gold arise in the case of the Netherlands
Less than 10 percent of total reserves, o
the average, are held as dollars. Since th
dollars are largely in the form of Treasur
bills a check was made to see if Treasur
bills held relative to gold were sensitive t
the return on dollars holding the Euro-dollar rate constant. This proved to be th
case with a fall in B/GA as theEuro-dolla
rate rose and a rise in B/GA as a weighte
average of the Treasury bill rate rose. Inclusion of the expected return on gol
yields the wrong sign. Apparently the absolutely and relatively small amount of it
total dollar assets leads the Netherland
to view dollar balances as a separate portfolio from gold. Behavior with regard t
gold indicates evidence of recognized interdependence.

The United Kingdom appears to behav
perversely regarding its actions relative t
the expected return on gold. Here, the reason is simple. While in most cases Swap
have been used by the Federal Reserv
System to gain exchange from its partners
the United Kingdom has made extensiv
use of its Swap facilities with the Federa
Reserve System to get dollars from th
United States in times of crises (over \$4
billion from 1962–67). Since a currenc
crisis generally raises the expected retur
on gold and the use of Swap facilities b
the United Kingdom has coincided wit
such crises, the data show a rise in dolla

---

deposits) is only a small portion, (an average of about 6 percent) of total dollar assets for Germany, Italy, and Switzerland. Since management of such a small portion of dollar assets as a part of a portfolio with gold reduces the possibility of a crisis, (say a rise in R_G leads to an exchange of dollars for gold), such behavior may be taken as evidence of a degree of recognized interdependence.

assets when the expected return on gold rises. It may be added that as a reserve currency country itself, the United Kingdom would stand to lose heavily if the United States had to devalue. That is, evidence of recognition by the United Kingdom of the interdependent nature of its actions is not surprising.

In order to consider the impact of an independent variable which might shift the M_{Li} (marginal return on dollars) schedule, the rate of change of prices in the United States[17] and the *U.S.* liquidity deficit,[18] were included separately in all of the equations run. Neither was significant in many cases. Either is included when significant (Table 1). It is worth noting that in a number of cases dollars held relative to gold simply rose with the *U.S.* liquidity deficit indicating a role for the United States as a viable international financial intermediary.

As it has been claimed[19] that in December and June Switzerland and Germany cut foreign exchange holdings and add to gold holdings as "window-dressing," a dummy variable for these months and countries was included. The results indicated the reverse though at a barely significant and in some cases, insignificant, level. For this reason they are not included in the results presented.

The $\bar{R}^2$ for Germany in Table 1 is markedly below that for most of the countries reported. The proximate reason for this is the low significance of the lagged dependent variable. This implies that for Germany the partial adjustment hypothesis may not be a correct specification of the portfolio adjustment process. In order to check for this possibility the equation was fitted omitting the lagged dependent variable but maintaining the modified first-differencing format given by equation (9) in order to adjust for possible autocorrelation. The result gave, for Germany:

$$(D/GA_t - \theta D/GA_{t-1})$$
$$= \underset{3.00}{.03108} - \underset{2.86}{.007634}(R_{GW_t} - \theta R_{GW_{t-1}})$$
$$- \underset{4.32}{.005056}(QGW_t - \theta QGW_{t-1})$$
$$+ \underset{1.31}{.005729}(R_{EDW_t} - \theta R_{EDW_{t-1}})$$
$$\bar{R}^2 = .30 \qquad F = 11.01 \qquad DW = 1.80$$

The estimated coefficients are not distinctly different and the fall in $\bar{R}^2$ is not large. It appears that the notion of partial adjustment is a weak hypothesis in the case of Germany.

The procedure just described in the case of Germany was performed for all countries included in Table 1. At the same time an attempt was made to see if seasonal dummy variables would prove significant. In only a few cases did seasonal factors prove significant. The countries for which the results do not permit rejection of the hypothesis of sensitivity of dollar and gold holdings to (expected) rates of return remain largely the same. (Results of fitting these equations are available from the author on request.)

Returning to Table 1, empirically the impact of the variance and skewness of returns upon portfolio behavior of the central banks was not striking. Allowing for heavier weighting of older values for the skewness of the return on gold, the calculated coefficient of skewness was significant with the correct sign at the .01 level in only three cases:[20] Germany, Italy and Sweden.

---

[17] Relative to the rate of change of prices in the country in question where available.

[18] The loss of *U.S.* gold plus the rise in *U.S.* short-term liabilities abroad.

[19] See Charles P. Kindleberger.

[20] Out of the total of thirteen countries either variance, skewness, or the correlation coefficient were significant in only five of the final equations. Discussion of these variables concerns only these cases.

The results for skewness need to be discussed along with those on variance. In no case did the variance of the return on gold prove to be a significant variable in explaining the ratio of any portion of dollar assets to gold. The cubic utility function referred to in Section I implies that the impact of variance upon the expected utility from an asset varies with the level of returns R. Specifically, behavior changes from risk-averse to risk-taking as R rises given $U'(R) > 0$ at all levels of R. If the same country wavered between risk-averse and risk-taking behavior (in terms of the cubic utility function), the coefficient associated with the variance would be likely to display no consistent value over time. No such explanation, however, can serve for the skewness and although it performs better than variance in explaining portfolio behavior of some central banks, it is not significant in a sufficient number of cases to be called an important independent variable.

In considering the results for the correlation coefficient between the expected return on gold and the return on dollars included as an independent variable in cases where variance and skewness appear, recall that a dollar asset/gold ratio of less than unity implies a fall in that ratio as the correlation between the return on gold and dollars rises. The ratio was less than unity for all cases save Sweden where the sign associated with ρ_{EG}, although incorrect, is not significant. Only in the case of Spain does ρ_{EG} over time perform at a significant (.01) level with the correct sign in determining the choice between gold and dollars. In four of five cases, countries appear to ignore ρ_{EG} in portfolio decisions.

It appears, then, that after adjustment for possible specification errors, the only independent variables with some statistically significant influence on the portfolio behavior of nine out of thirteen major central banks are the return on dollar assets and the expected return on gold. The skew-
ness of the expected return on gold is st
tistically significant in three of thirtee
cases. These results arise when dollar-a
set holdings and gold holdings are adjuste
for Swaps and Roosa Bonds where such a
in use. In some cases, for particularly lar
central banks like Germany, Italy ar
Switzerland, portfolio behavior can be i
ferred for only a small portion of doll
assets D relative to gold.

E. Implications of the Results for Long-Run Portfolio Adjustment

Before proceeding to an assessment
the results taken as a whole, it is importa
to consider the implications of the stati
tically significant results presented
Table 1 in terms of longer-run behavio
This requires some consideration of t
manner in which these coefficients we
estimated.

Recall that since the observations of t
study are monthly, a partial adjustme
process was assumed.[21] Estimation of t
equations including the dependent va
able lagged one period as an independe
variable provided a consistent estimat
of λ the partial adjustment coefficient. T
coefficients estimated for the independe
variables are then "short-run" coefficien
since the true coefficients are all multipli
by λ. The "long-run" or full coefficien
may then be estimated from the equatio
actually fitted by dividing the short-r
coefficients by λ. This is done in Table
for the coefficients associated with t
expected returns on gold and with t
return on dollars.

The mean results indicate that for t

---

[21] The implicit assumption is that the cost of compl
ing full adjustment within a month is too high. Esti
tion of the equations assuming this to be the case yi
an estimate of $\lambda\beta_i$ and of $(1-\lambda)$. A consistent estima
of λ can be obtained from $1-(1-\lambda)$. Then, a consiste
estimator of $\beta_i = \tilde{\beta}_i$ can be obtained from the expressi
$(\lambda\beta_i/\tilde{\lambda}) = \tilde{\beta}_i$, where $\tilde{\beta}_i$ is not BLUE but it is consiste
and asymptotically efficient as the sample size
proaches infinity. It should be noted that $\tilde{\beta}_i$ obtain
from Table 2 is an estimate from eighty-one obser
tions through time and a cross section of nine countri

TABLE 2—ESTIMATION OF LONG-RUN COEFFICIENTS FOR THE EXPECTED RETURN ON GOLD AND THE RETURN ON DOLLARS, USING $\tilde{\lambda}$

Country	$\lambda\beta_1$[a]	$\lambda\beta_4$[a]	$\tilde{\lambda}$[b]	$\tilde{\beta}_1=\lambda\beta_1{}^b/\tilde{\lambda}$	$\tilde{\beta}_4=\lambda\beta_4{}^b/\tilde{\lambda}$
France:	−.02172	−.00688	+.30840	−.07040	−.02230
	4.12	*1.02*	*5.11*	*5.68*	*1.05*
Germany:	−.00709	+.00558	+.90360	−.00785	+.00617
	2.67	*1.25*	*9.31*	*2.76*	*1.26*
Italy:	−.01027	+.01626	+.22030	−.04660	+.07381
	3.74	*3.44*	*3.72*	*3.17*	*2.99*
Switzerland:	−.00527	+.00637	+.25000	−.02000	+.02400
	2.34	*1.88*	*2.98*	*2.80*	*1.31*
Denmark:	−.02936	+.06663	+.13490	−.21760	+.49390
	2.38	*2.89*	*2.31*	*1.60*	*2.44*
Sweden:	−.05273	+.11120	+.07100	−.74270	+1.5660
	2.02	*2.30*	*1.55*	*1.42*	*2.16*
Spain:	−.01091	+.00828	+.13510	−.08075	+.06129
	2.12	*1.01*	*2.60*	*2.02*	*0.87*
Belgium:	−.00313	+.01238	+.37660	−.00830	+.03287
	2.40	*3.97*	*4.80*	*2.44*	*5.50*
Japan:	−.08827	+.12300	+.24310	−.36310	+.50600
	3.43	*2.53*	*3.79*	*3.17*	*2.25*
Mean:	−.02541	+.03090	+.29370	−.17300	+.30500
Weighted Mean:[c]				−.01900	+.11900

[a] Taken from Table 1. Figures in italics are *t*-ratios.

[b] The italicized figures are the coefficient estimates divided by the standard errors. The standard errors are obtained from consistent estimators of the variance and covariance of $\tilde{\lambda}$, $\tilde{\beta}_1$ and $\tilde{\beta}_4$. For a description of the method used to obtain these estimates, see A. S. Goldberger et al.

[c] See fn. 23 for the weighting scheme used in computing the weighted means.

nine countries[22] considered, the *ratio* of dollars to gold when adjustment was complete, was estimated to be more sensitive, in absolute terms, to the return on dollars than it was to the return on gold. (Table 2, col. 4 & 5, "mean" and "weighted mean.")

Since the countries in the sample are by no means of equal significance in terms of the size of their holdings of dollars and gold, the relevant coefficients of complete adjustment (long run) are computed as a weighted[23] sum of the coefficients for each country. The estimates are as follows:

$$\frac{\partial}{\partial R_L}(L/G) = +.119; \quad \frac{\partial}{\partial R_G}(L/G) = -.091$$

It can easily be shown[24] that these re-

---

[22] The other four major countries considered did not yield decisive enough results to be included (See Table 1).

[23] Weight of country i = (average of gold plus dollar holdings 1961–68 for country i/average of gold plus dollar holdings 1961–68 for all nine countries).

$$W_i = \left[\sum_{t=1}^{8}(G_t + D_t) \Big/ \sum_{t=1}^{8}\sum_{i=1}^{9}(G_{ti} + D_{ti})\right],$$

(where i = countries and t = years).

[24]
$$\begin{aligned} \partial/\partial R_L(L/G) &= +0.119 \\ &= (1/G)\partial L/\partial R_L - (L/G^2)\partial G/\partial R_L \\ \partial/\partial R_G(L/G) &= -0.091 \\ &= (1/G)\partial L/\partial R_G - (L/G^2)\partial G/\partial R_G \end{aligned}$$

Substituting in the average values of L and G for the sults imply a value of $\partial L/\partial R_L > \partial G/\partial R_G$. I hypothesize that a reason for the lower sensitivity of gold holdings to changes in the return on gold relative to sensitivity of dollar holdings to changes in the return on dollars is the higher cost of making portfolio adjustments in gold. Dollars could be obtained from any central bank which maintains convertibility between its currency and dollars. Gold must be purchased in London or from the U.S. Treasury. The Treasury gives up gold reluctantly and may impose some costs on countries which persist in asking the Treasury to honor its promise on the dollar price of gold. Or the Treasury may try to sell Roosa Bonds as a gold substitute bearing interest. All of this is likely to entail costly negotiations. Purchases in London, if excessive, could bring the gold pool into operation (before March 1968) and would eventually bring a gold pool country wishing to raise the gold portion of its reserves back to the U.S. Treasury. This result confirms the prediction of the model for a case where transaction costs are higher for one asset than for the other.[25]

III. Summary and Interpretation of Empirical Results

Recognition of the inhomogeneity of dollar assets and modification of independent variables to allow for different distributed lags, produced portfolio-type behavior i nine of the thirteen cases considered. T the degree that only some of the dolla held by some of the major central bank were managed as a part of a portfolio wit gold, the assumption of independent uti ity-maximizing behavior must be que tioned. The conclusion that can still b drawn remains: some part of dollar asse of some major central banks is manage in a manner which indicate they are viewe as part of a portfolio with gold. Furthe more, conversion of short-run results int long-run terms yields results consiste with the reasonable hypothesis that tran actions costs are higher for gold than f dollars.

All of the implications of the cubic uti ity function for portfolio management reserves are not, however, borne out f large central banks. This conclusion d rives from the generally low significance variance and skewness of the expecte return on gold and/or lack of significan of the correlation coefficient with the r turn on dollars in explaining holdings dollars relative to gold.

It is important to remember that th impact of adjustment for both Swaps an Roosa Bonds bears on the results for Ge many, Italy, Switzerland, Austria, an Belgium, while adjustment for Swaps bea on results for France, the United Kingdo and the Netherlands. The amount Swaps actually in use, as noted above, subtracted from dollar assets along wit Roosa Bonds since they are treated equivalent to a conversion of dollars int gold. These facilities have been show elsewhere[26] to be systematically related t the expected return on gold, its varian and skewness. The question arises as t whether such arrangements are of an i dependent or an interdependent natur To the extent that they require direct n

---

nine countries over the 1961–68 period gives:

$$+.119 = 0.063\partial L/\partial R_L - 0.043\partial G/\partial R_L$$

$$-.091 = 0.063\partial L/\partial R_G - 0.043\partial G/\partial R_G$$

This gives $\partial L/\partial R_L > \partial G/\partial R_G$ for any assumed ratio of cross partials to own partials.

[25] This manner of analyzing the choice among assets is derived from the treatment of the topic by Sir John Hicks. See ch. 6 especially. His analysis also includes a means to consider the impact of transactions costs on portfolio adjustment. Inclusion of such costs in the analysis would produce a kink in the M_1 and M_2 lines at the point representing the level of initial holdings. The sharpness of the kink is directly proportional to the level of transactions costs. A sharper kink for asset G than for asset L indicates a lesser shift into G, given a rise in R_G, than into L given a rise in R_L.

[26] See Makin (1971).

gotiations between central banks with the U.S. Federal Reserve for Swaps and with the U.S. Treasury for Roosa Bonds, they manifest cooperation which implies some recognition of interdependence. Yet to some extent they manifest independence, because they offer a way to adjust the portfolio in terms of risk. These facilities are, I believe, symptomatic of the blend of independence and uneasy interdependence which was characteristic of the international monetary system in the 1960's.

IV. Postscript

As this article was in its final stages of preparation, the United States closed the gold window, formally ending for foreign central banks, the convertibility of dollars into gold at a fixed price. This move was largely dictated by a desire of the United States to prevent foreign central banks from exercising control over its own monetary policy by threatening to exchange some of their rapidly expanding supply of dollars for the dwindling *U. S.* gold stock.

This article suggests that central banks are indeed not indifferent to the composition of their international reserves. This fact, along with the relative constancy of official gold stocks which has accompanied the two-tier gold market, suggests that if total international reserves are to grow by way of increased holdings of some national currency, a confidence problem will almost inevitably arise. The crisis of confidence may not be so much over the *de facto* convertibility of a national currency into a commodity like gold. Rather, it is likely to be over a fear that the rate of growth of international reserves, implicit in the monetary policy of the country whose currency is used as international reserves, will contradict or complicate domestic economic management for countries outside of the reserve currency country. Convertibility of the reserve currency into some commodity like gold at a fixed price assures other countries of an ability to restrict the rate of monetary expansion in the reserve currency country. Such an ability produces a basis for conflict between domestic policy goals and the international role of its money within the reserve currency country.

In view of these problems, which hopefully this paper has helped to demonstrate are likely to occur, the first priority for international monetary reform ought to be removal of a national currency from a key role in the international monetary system. SDR's exist as a viable means to accomplish this goal. It would also be desirable to cut the link between SDR's and a single commodity like gold. Management of their quantity by IMF members to insure stable prices defines a link between the quantity of SDR's and the quantity of all goods and services entering into world trade.

REFERENCES

S. Almon, "The Distributed Lag Between Capital Appropriations and Expenditures," *Econometrica,* Jan. 1965, *33,* 178–96.

P. W. Bell, "Private Capital Movements and the U.S. Balance of Payments Position," in *Factors Affecting the U.S. Balance of Payments,* Joint Economic Committee, 87th Cong., 2d sess., 1962, Washington, pp. 395–481.

J. Durbin, "Testing for Serial Correlation in Least Squares Regression when some of the Regressors are Lagged Dependent Variables," *Econometrica,* May 1970, *38,* 410–21.

D. E. Farr, *The Investment Decision under Uncertainty,* Englewood Cliffs, 1962.

Z. Griliches, "Distributed Lags: A Survey," *Econometricia,* Jan. 1967, *35,* 16–49.

A. S. Goldberger, A. C. Nagar, and H. S. Odeh, "The Covariance Matrices of Reduced-Form Coefficients and of Forecasts for a Structural Econometric Model," *Econometrica,* Oct. 1961, *39,* 556–73.

H. A. Hagemann, "Reserve Policies of Central Banks and their Implications for U.S. Balance of Payments Policy," *Amer. Econ. Rev.*, Mar. 1969, *59*, 62–77.

G. Hanoch and H. Levy, "Efficient Portfolio Selection with Quadratic and Cubic Utility," *J. Bus., Univ. Chicago*, Apr. 1970, *43*, 181–89.

J. R. Hicks, *Critical Essays in Monetary Theory*, Oxford 1967.

P. B. Kenen, "International Liquidity and the Balance of Payments of a Reserve Currency Country," *Quart. J. Econ.*, Nov. 1969, *74*, 572–86.

C. P. Kindleberger, *International Economics*, Homewood 1968.

H. Levy, "A Utility Function Depending on the First Three Moments," *J. Finance*, Sept. 1969, *24*, 715–19.

F. Machlup, *Remaking the International Monetary System: The Rio Agreement and Beyond*, Baltimore 1968.

G. S. Maddala, "A Note on Jorgenson's and Almon's Distributed Lag Model," mimeo. 1969.

J. H. Makin, "Reserve Policies of Central Banks: Comment," *Amer. Econ. Rev.*, Mar. 1970, *60*, 210–11.

———, "Swaps and Roosa Bonds as an Inde of the Cost of Cooperation in the Crisi Zone," *Quart. J. Econ.*, May 1971, *85*, 349–56.

E. Malinvaud, *Statistical Methods of Econometrics*, Amsterdam 1966.

M. Nerlove and K. F. Wallis, "Use of th Durbin-Watson Statistic in Inappropriat Situations," *Econometrica*, Jan. 1966, *34* 235–38.

L. H. Officer and T. D. Willett, "Reserve Asset Preferences and the Confidence Proble in the Crisis Zone," *Quart. J. Econ.*, Nov 1969, *83*, 688–95.

F. Pick, *Pick's Currency Yearbook*, New Yor 1961–62/1967–68.

R. Triffin, *Gold and the Dollar Crisis*, Ne Haven 1960.

Federal Reserve Bank, *Fed. Reserve Bull.*, vols 49–54, Mar.-Sept., Washington 1963–68.

International Monetary Fund, *Internationa Financial Statistics*, vols. 16–21, Washingto 1963–68.

Organization for Economic Cooperation an Development, *Main Economic Indicators* Paris 1963–68.

Investment in Graduate Education

By YORAM WEISS[*]

The wealth of a nation is greatly influenced by the quality of its labor force, and this in turn depends upon the propensity of each individual to make the best use of his abilities and to improve himself through learning. To induce changes in these individual "investment" decisions is a basic target of economic policy. In this respect, the empirical and theoretical investigations of the role of monetary motives in the choice of a "life plan" should prove highly relevant.

In this paper I examine the educational and occupational choices of graduate students in the natural and social sciences. Presumably, the type and length of graduate education are subject to choice. My working hypothesis is that in making these choices, students attempt to maximize the present value of lifetime earnings. The main purpose is to test whether this hypothesis is, in fact, consistent with the actual choices of scientists.

In the first part of this paper, I restate briefly the conventional theory of investment in human capital and its empirical implications in a certain world with perfect capital markets.

In the second part, I derive estimates of the wage functions in different scientific fields. These estimates are then utilized to calculate the rates of return and the lifetime earning streams which are associated with different types and levels of graduate education.

The last part of the paper is an attempt to relate the observed choices of graduate students to lifetime earning differentials across fields.

I.

The Wage Function Hypothesis

An individual with a given ability endowment can affect his earning power by a suitable choice of occupation, level of education, and degree of specialization.

Suppose that n occupations are considered by the individual. Let w_i be the wage rate he expects in occupation i, t his age, δ_i the exogenous rate of change in his earning power in occupation i, A the ability endowment of the individual, H_i his accumulated work experience in occupation i, and E_i the accumulated amount of schooling specific to occupation i. We may then form the wage function of the individual as

$$w_i = e^{\delta_i t} f_i(A, H_i, E_i) \tag{1}$$

The wage function (1) describes the wage structure which the individual expects. It is derived essentially from technological considerations, but it also depends on the state of demand and supply for each grade of labor. The basic elements of the wage function hypothesis seem self-evident. Employers are willing to offer higher wages to the able, educated, and experienced worker, who is presumably capable of doing anything the less able, educated, and experienced worker can do, and a little more. In other words, able, experienced, and educated employees are more "productive," and in relatively limited supply. We assume therefore that the first-order derivatives of the wage function are all positive.

---

[*] Department of economics, Tel-Aviv University, Israel. This paper is based on a Ph.D. dissertation submitted to Stanford University in August 1968. I am most grateful to M. Reder who was my main advisor; I wish to thank J. Pencavel, J. Rosse, Y. Mundlak, G. Fishelzon and R. Hollister for their comments. Special thanks are due to M. Levine of the National Science Foundation for providing the data and aiding in its interpretation.

We represent the net outcome of the various exogenous changes that the individual expects in the future by δ_i. He may expect certain shifts in demand which would cause wages to rise in some occupations and to decline in others. His earning power is likely to decay due to the natural processes of aging. The value of his human capital stocks may depreciate, as part of what is learned is forgotten, and old vintages become obsolete due to technological improvements in schooling. For simplicity, we assume a constant rate of change (which may be positive or negative) at each occupation.

Note the following restrictive characteristics of the wage function hypothesis. Both experience and education are considered here to be completely specific to an occupation. Ability, on the other hand, is treated as perfectly general. For the problem at hand, this assumption is not unreasonable. In contrast to lower levels of education which tend to provide general knowledge, graduate studies are mostly specialized. Experience is also likely to be specific to a scientific field when a broad classification is adopted. Finally, a similar type of ability may be required in all scientific fields.

It is assumed that both experience and education can be measured by their accumulated duration. No allowance is made for differences in the quality of schools or jobs in providing education and experience.

In a static world, or alternatively, in a balanced growth situation where $\delta_i = \delta$, the wage function hypothesis implies that a maximizing individual will concentrate his investment in schooling as early as possible, and specialize in *one* occupation thereafter.[1] We shall assume throughout that this is in fact the typical pattern. We may then describe the individual as following a two-step decision process. First, he determines the optimal amount of education for any given occupation, then h chooses the occupation which, with its optimal amount of education, yields the highest present value of lifetime earnings.

By assumption, the maximizing indivi ual spends all his time on productive tim activities (work or schooling). It thus fo lows from occupational specialization tha

$$H_i + E_i = \mathrm{t} \tag{2}$$

It is therefore possible to eliminate H_i an to view the wage function as an explici function of age, $w_i(\mathrm{t}) = e^{\delta_i \mathrm{t}} f_i(A, \mathrm{t} - E_i, E_i)$. I $\delta_i < 0$ and if there is decreasing margin productivity of experience, then for a give level of education, $w_i(\mathrm{t})$ will be a singl peaked function of age;[2] if $\delta_i > 0$ the earnin profile will be monotonically increasing.

The individual may have some earning grants, or direct costs during his schoolin period. Let us write the "students' ne earning function" as $s_i = s_i(A, E_i)$. We ma assume that the more able and educate student enjoys higher net earnings. It likely that at high levels of education (e.g graduate school) net student earnings a positive.

The Choice of the Optimal Life-Program

The present value of lifetime earning enjoyed by an individual who plans to a quire education of type i, and work in o cupation i thereafter is

$$\begin{aligned} V_i(A, E_i, r, \delta_i) &= \int_0^{E_i} e^{-r\mathrm{t}} s_i(A, \mathrm{t}) d\mathrm{t} \\ &\quad + \int_{E_i}^{\mathrm{T}} e^{-(r-\delta_i)\mathrm{t}} f_i(A, \mathrm{t} - E_i, E_i) d \end{aligned} \tag{3}$$

---

[1] See Weiss (1968, ch. 1) and (1971a).

[2] The peak is determined by the condition

$$\frac{dw_i}{d\mathrm{t}} = e^{\delta_i \mathrm{t}} \left(\frac{\partial f_i}{\partial H_i} + \delta_i f_i \right) = 0, \qquad \delta_i < 0$$

The second derivative at this point is

$$e^{\delta_i \mathrm{t}} \left(\frac{\partial^2 f_i}{\partial H_i^2} + \delta_i \frac{\partial f_i}{\partial H_i} \right) < 0;$$

there is thus only one peak.

$$(4)\quad \frac{\partial V_i}{\partial E_i} = e^{-(r-\delta_i)E_i}\left\{e^{-\delta_i E_i}s_i(A,E_i) - f_i(A,O,E_i) + \int_{E_i}^{\mathrm{T}} e^{-(r-\delta_i)(\mathrm{t}-E_i)}\left[\frac{\partial f_i(A,\mathrm{t}-E_i,E_i)}{\partial E_i} - \frac{\partial f_i(A,\mathrm{t}-E_i,E_i)}{\partial H_i}\right]d\mathrm{t}\right\} = 0$$

$$(4')\quad \frac{\partial V_i}{\partial E_i} = e^{-(r-\delta_i)E_i}\left\{e^{-\delta_i E_i}s_i(A,E_i) - e^{-(r-\delta_i)(\mathrm{T}-E_i)}f_i(A,\mathrm{T}-E_i,E_i) + \int_{0}^{\mathrm{T}-E_i} e^{-(r-\delta_i)\mathrm{t}}\left[\frac{\partial f_i(A,\mathrm{t},E_i)}{\partial E_i} - (r-\delta_i)f_i(A,\mathrm{t},E_i)\right]d\mathrm{t}\right\} = 0$$

where r is the market rate of interest (expected to remain constant) and T is the length of the individual's life.

The optimal amount of education E_i^* is determined from the maximization of V_i with respect to E_i. The first-order condition (for interior optimum) is shown in equation (4) and in alternative form in (4′). That is, the individual will continue to acquire education up to the point where his earning power (opportunity cost) is equal to the present value of the expected additional earnings resulting from additional education (and foregone experience). Alternatively, the costs resulting from the loss of the last year of experience and delayed earnings must be equal to the gain due to increased earning power.

Clearly, an interior solution may not exist. If, for instance, students' earnings are always above the earning capacity out of school, the individual will never stop studying. Similarly, if the marginal productivity of experience is always above the marginal productivity of schooling, the individual will take no schooling at all. We shall assume that neither of these happens so that for at least some rates of interest there is an interior solution. (More specifically, we assume $\partial w_i/\partial E_i > \partial w_i/\partial H_i$, and $s_i\,(A,E_i) < e^{\delta_i t}f_i(A,0,E_i)$.)

It is convenient to state the first-order condition in a somewhat different form.

Let us define:

$$(5)\quad g_i(A,E_i,\rho,\delta_i) = e^{-(\rho-\delta_i)E_i}\left\{e^{-\delta_i E_i}s_i(A,E_i) - f_i(A,0,E_i) + \int_{E_i}^{\mathrm{T}} e^{-(\rho-\delta_i)(\mathrm{t}-E_i)}\left[\frac{\partial f_i}{\partial E_i} - \frac{\partial f_i}{\partial H_i}\right]d\mathrm{t}\right\}$$

where ρ is a real number. Clearly, $g_i(A,E_i,r,\delta_i) \equiv \partial V_i/\partial E_i$.

Now consider the condition:

$$(6)\qquad g_i(A,E_i,\rho,\delta_i) = 0$$

For a given A and δ_i, condition (6) defines a unique ρ for every level of education E_i.[3] We shall call this function the *marginal rate* of return to education in occupation i.[4] An alternative formulation of the first-order condition is condition (7).

$$(7)\qquad \rho_i(A,E_i,\delta_i) = r$$

That is, the optimal level of education is that which equates the marginal rate of return to the market interest rate.

The second-order condition for achieving

[3] At the point where $g_i(\ ) = 0$,

$$\frac{\partial g_i}{\partial \rho} = -\int_{E_i}^{\mathrm{T}} e^{-(\rho-\delta_i)\mathrm{t}}(\mathrm{t}-E_i)\left(\frac{\partial f_i}{\partial E_i} - \frac{\partial f_i}{\partial H_i}\right)d\mathrm{t} < 0$$

there is thus at most one ρ where $g_i(\ ) = 0$.

[4] The usage of the terms "marginal" and "average" rates of return and the analysis in this section are based on Friedrich and Vera Lutz, ch. 1.

a maximum is

$$(8)\quad \frac{\partial^2 V_i}{\partial E_i^2} < 0 \text{ or equivalently}^5\ \frac{\partial \rho_i}{\partial E_i} < 0,$$

that is, decreasing marginal rate of return to education.[6]

In contrast with the marginal rate of return which serves to determine the optimal level of education, the average rate of return can be used to determine whether or not investment in education of a given size should be undertaken. We may define the profits from a given level of education as

$$(9)\quad \begin{aligned} P_i(A, E_i, r, \delta_i) &= V_i(A, E_i, r, \delta_i) \\ &\quad - V_i(A, 0, r, \delta_i) \end{aligned}$$

Now consider the function $P_i(A,E,\gamma,\delta_i)$ where γ is a real number. The condition of zero profits defines a unique[7] γ for every level of education E_i when A and δ_i ar given. This function is the *average rate* o return from education at occupation i. Th marginal and average rates of return ar related in the familiar way. That is, th average rate of return is decreasing if th average rate exceeds the marginal rate an increasing if the average rate is below th marginal rate.[8]

The conditions (7) and (8) are necessar and sufficient for local maximum. An additional necessary condition for globa maximum is that the average rate of retur must be above the rate of interest, otherwise the individual would be better of with no education at all.

We may use an occupation which is ope to any college graduate as a standard o comparison. Let $V_0(A,0,r,\delta_0)$ be the presen value of lifetime earnings in this occupation; then the profits from occupation would be

$$(10)\quad \begin{aligned} \Pi_i(A, r, \delta_i) &= V_i(A, E_i^*(A, r, \delta_i), r, \delta_i) \\ &\quad - V_0(A, 0, r, \delta_0) \end{aligned}$$

and the average rate of return from occupation i, or alternatively of education o type i, would be that level of r at whic $\Pi_i = 0$.

The choice of occupation is determine from the maximization of Π_i over all possible i.

We thus have two basic hypotheses: 1 the wage function hypothesis; and 2) th

---

[5] From (6) we obtain

$$\frac{\partial g_i}{\partial E_i} dE_i + \frac{\partial g_i}{\partial \rho} d\rho = 0$$

but $\partial g_i/\partial \rho < 0$ hence $d\rho/dE_i < 0$ if, and only if, $\partial g_i/\partial E_i < 0$. Since the first-order condition is satisfied, $r = \rho$ and $\partial g_i/\partial E_i = \partial^2 V_i/\partial E_i^2$. Therefore, $d\rho/dE_i < 0$ if, and only if, $\partial^2 V_i/\partial E_i^2 < 0$.

[6] At the optimum

$$\begin{aligned} \frac{\partial^2 V_i}{\partial E_i^2} = e^{-(r-\delta_i)E_i} \Bigg\{ &-e^{-(r-\delta_i)(\mathrm{T}-E_i)} \left[2\frac{\partial f_i}{\partial E_i} - \frac{\partial f_i}{\partial H_i} \right] \\ &+ \int_0^{\mathrm{T}-E_i} e^{-(r-\delta_i)t} \left[\frac{\partial^2 f_i}{\partial E_i^2} - (r-\delta_i) \frac{\partial f_i}{\partial E_i} \right] dt \\ &+ e^{-\delta_i E_i} \left[\frac{\partial s_i}{\partial E_i} - \delta_i s_i \right) \Bigg\} \end{aligned}$$

The first term is negative by assumption; the second term will be negative if we assume decreasing marginal productivity of education; the whole expression will be negative if $\partial s_i/\partial E_i$ is "small."

[7] $P(A, E, \gamma, \delta_i)$ is not monotone in γ but to show uniqueness, it is sufficient to prove that $\partial P_i/\partial \gamma < 0$ whenever $P \geq 0$.

$$\begin{aligned} \frac{\partial P_i}{\partial \gamma} = &\int_0^{E_i} \mathrm{t}e^{-(\gamma-\delta_i)\mathrm{t}}[f_i(A, \mathrm{t}, 0) - e^{-\delta_i \mathrm{t}} S_i(A, \mathrm{t})]d\mathrm{t} \\ &- \int_{E_i}^{\mathrm{T}} \mathrm{t}e^{-(\gamma-\delta_i)\mathrm{t}}[f_i(A, \mathrm{t} - E_i, E_i) - f_i(A, \mathrm{t}, 0)]d\mathrm{t} \end{aligned}$$

since by assumption $e^{\delta_i \mathrm{t}} f_i > S_i$ and $f_i(A, \mathrm{t} - E_i, E_i) > f_i(A, \mathrm{t}, 0)$

$$\begin{aligned} \frac{\partial P_i}{\partial \gamma} < \int_0^{E_i} E_i e^{-(\gamma-\delta_i)\mathrm{t}}[f_i - e^{-\delta_i \mathrm{t}} S_i]d\mathrm{t} - \int_{E_i}^{\mathrm{T}} E_i e^{-(\gamma-\delta_i} \\ [f_i(A, t - E_i, E_i) - f_i(A, t, 0)]d\mathrm{t} \end{aligned}$$

Thus $\partial P_i/\partial \gamma < -E_i P_i(A, E_i, \gamma, \delta_i)$.

[8] $$\frac{d\gamma}{dE_i} = -\frac{\partial P_i}{\partial E_i} \bigg/ \frac{\partial P_i}{\partial \gamma} = -g_i(A, E_i, \gamma, \delta_i) \bigg/ \frac{\partial P_i}{\partial \gamma}$$

since $P_i = 0$, $\partial P_i/\partial \gamma < 0$. If $\gamma = \rho$ then $g_i = 0$. If $\gamma > \rho$ the $g_i < 0$ (recall that $g' < 0$) and $(d\gamma/dE_i) < 0$, and vic versa if $\gamma < \rho$. As noted by Lutz and Lutz, "The relation ship between the marginal and average rate of retur is *not* . . . exactly similar to that between margina and average revenue (or marginal and average costs) (p. 20).

maximization of the present value of lifetime earnings hypothesis. Implicitly we assume perfect foresight and perfect capital markets. There are two sets of testable implications: one set follows from the "individual experiment" while the other set is a result of a "market experiment."

On the individual level, we have the following implications: Investment in schooling will be concentrated at an early age. The degree of occupational mobility among broadly classified fields will be small. All individuals will attain the same marginal rate of return from education. A higher rate of interest will discourage investment in schooling. Individuals who are equally able will invest the same amount of time in education and choose the same occupation. The occupation which is considered best for such an "ability class" may vary with the interest rate. An able individual may choose to invest less than a less able one in education[9] but irrespective of his choice he will always enjoy higher lifetime earnings. (He can always make the same choice and earn more.)

At the market level, the following implications hold: In a state of long-run equilibrium, an individual of given ability will be indifferent among all alternative programs. He will have the same lifetime earnings in all occupations, provided that he is choosing the optimal amount of education for each occupation. The average rates of return from all occupations will be equal, but the average rate of return from education within each occupation may differ. If a state of disequilibrium exists and some ability class finds a certain occupation more profitable than others, a flow of new entrants will push down the earnings of all workers in this occupation. The costs of acquiring the corresponding type of education will increase. Eventually, equilibrium will be restored. Adjustment, however, may be very slow. Individuals who are already in the labor force are not likely to adjust, in a state of disequilibrium, since the accumulation of experience, which is to a large extent occupation specific, imposes considerable costs. The burden of adjustment is almost exclusively on new entrants.

II.

The highly simplified model which is presented in the first part of the paper may be used for the analysis of the observed choices of graduate students in the natural and social sciences.

My main source of data is a random sample of 5686 scientists from the 242,800 who reported to the National Register of Scientific and Technical Personnel in 1966.[10] The scientists in the sample were classified into eight fields: chemistry, physics, earth science, mathematics, biology, agriculture, psychology, and the social sciences (economics, sociology and statistics). Following the practice used in the National Register, the scientist's field or occupation is the one in which he reports to be the most competent on the basis of his total educational and work experience.

The Observed Lifetime Pattern of the Typical Scientist

Let us start with some observations on the timing of the investment in education

---

[9] From the first-order condition $g_i(A, E_i, r, \delta_i)=0$ we obtain

$$\frac{dE_i^*}{dA} = -\frac{\partial g_i}{\partial A} \Big/ \frac{\partial g_i}{\partial E_i}$$

From the second-order condition $\partial g_i/\partial E_i<0$. Thus the sign of dE^*_i/dA is equal to the sign of $\partial g_i/gA$.

$$\frac{\partial g_i}{\partial A} = e^{rE_i}\frac{\partial S_i}{\partial A} - e^{-(r-\delta_i)E_i}\frac{\partial f_i}{\partial A} + \int_{E_i}^{T} e^{-(r-\delta_i)t}\left(\frac{\partial^2 f_i}{\partial E_i \partial A} - \frac{\partial^2 f_i}{\partial H_i \partial A}\right) dt$$

All the first- and second-order derivatives are likely to be positive, so that the sign of $\partial g_i/\partial A$ is undetermined. See also Weiss (1971b).

[10] For further information on this source, see "American Science Manpower," the 1964 report, pp. 1–3, and the 1966 report, pp. 3–6.

TABLE 1—CUMULATIVE DISTRIBUTION OF SCIENTISTS BY AGE AT WHICH DEGREE IS OBTAINED

Age	B.A.	M.A.	Ph.D.
21	12.1	1.1	0.2
22	44.2	4.7	0.6
23	60.0	14.7	1.8
24	68.9	28.4	5.6
25	76.4	40.6	12.9
26	82.0	50.1	23.9
27	86.9	58.7	35.3
28	89.8	66.1	44.8
29	92.6	72.5	54.1
30	94.1	77.6	62.3
31	95.4	81.9	69.3
32	96.3	84.5	74.8
33	96.9	87.1	79.9
34	97.5	89.0	84.0
35	97.7	90.9	87.9
36–45	99.7	97.8	97.9

and the degree of occupational specialization.

In Table 1 we present the cumulative distribution of scientists by age of obtained degree. If it is assumed that the average scientist who acquires his education continuously will obtain his B.A. degree at the age of 23, his M.A. degree at the age of 25 and his Ph.D. before he turns 30, it then seems that about 40 percent of the scientists with B.A. degrees, 60 percent of those holding M.A. degrees, and 30 percent of those with Ph.D. degrees delayed their investment in education. (Apparently, for shorter investment such as the M.A. program, more scientists can afford a delay.)

Uncertainty and imperfections in the capital market are the likely causes of the observed delay in the investment in education. The scientist may delay his schooling in order to ascertain the market situation and his own abilities and preferences. He may be unable to borrow on account of his future earnings and be thus forced to delay his education. Finally, the observed delay may be, in part, due to exogenous events, such as World War II which prevented some scientists from following their original plans.

Let us note, however, that the postponement of education is relatively short: very few scientists acquire their schooling at a mature age and about 90 percent complete their education before the age of 35.

In Tables 2 and 3 we present for each field the occupational distribution of scientists (Ph.D.'s in Table 3) who acquired their education in that field. As indicated by the dominance of the diagonal elements there is a strong tendency of scientists to remain in the field in which they have acquired their education. As one would expect, the rate of interoccupational mobility decreases with the level of education. In other words, higher levels of education tend to be more specific.

We may conclude that the working life pattern of the typical scientist is characterized by an early investment in education and occupational stability. The

TABLE 2—OCCUPATIONAL DISTRIBUTION BY FIELD OF STUDY—ALL SCIENTISTS

	Occupation								
Degree	Chem.	Physics	Earth Sciences	Math.	Bio.	Agr.	Psychology	Social Sciences	Others
Chemistry	0.77	0.04	0.01	0.00	0.05	0.04	0.00	0.03	0.06
Physics	0.02	0.82	0.04	0.05	0.01	0.00	0.00	0.00	0.07
Earth Sciences	0.01	0.01	0.93	0.01	0.00	0.00	0.00	0.00	0.03
Mathematics	0.01	0.02	0.03	0.80	0.00	0.00	0.00	0.06	0.08
Biology	0.03	0.00	0.01	0.01	0.85	0.05	0.01	0.00	0.04
Agriculture	0.01	0.00	0.01	0.00	0.11	0.85	0.00	0.02	0.01
Psychology	0.00	0.00	0.00	0.01	0.01	0.00	0.95	0.02	0.01
Social Sciences	0.03	0.00	0.01	0.10	0.01	0.00	0.02	0.72	0.10

TABLE 3—OCCUPATIONAL DISTRIBUTION BY FIELD OF STUDY—PH.D. HOLDERS

	Occupation								
Degree	Chem.	Physics	Earth Sciences	Math.	Bio	Agr.	Psychology	Social Sciences	Others
Chemistry	0.80	0.05	0.01	0.00	0.07	0.02	0.00	0.01	0.04
Physics	0.01	0.91	0.02	0.02	0.02	0.00	0.00	0.00	0.02
Earth Sciences	0.01	0.02	0.94	0.00	0.00	0.00	0.00	0.01	0.02
Mathematics	0.00	0.02	0.01	0.88	0.00	0.00	0.00	0.05	0.05
Biology	0.02	0.00	0.00	0.00	0.92	0.03	0.00	0.00	0.02
Agriculture	0.01	0.00	0.00	0.00	0.14	0.82	0.00	0.03	0.03
Psychology	0.00	0.00	0.00	0.01	0.01	0.00	0.94	0.03	0.02
Social Sciences	0.00	0.00	0.00	0.04	0.00	0.00	0.04	0.85	0.08

high degree of occupational stability enables us to derive estimates of the earning streams which are associated with alternative lifetime plans mainly from cross-section data. We may suppose that a person who is just about to make his choice forms his expectations on the basis of the currently observed earnings of older individuals, and that he modifies them according to his estimate of the general time-trend in the earnings at each occupation.

The Estimated Wage Function and the Rates of Return from Education

We now turn to the empirical estimates of the marginal and average rates of return for investment in higher education in the several scientific fields. The basic components which are necessary for the construction of these estimates are: 1) the expected lifetime earning streams associated with different levels of education, 2) the direct costs and income which accrue during the schooling period, 3) the amount of time needed to acquire each level of education. We shall consider each of these components in turn.

1) In each of the eight scientific fields, a wage function of the following form is estimated:

$$(11)\quad y_i = e_i^{\alpha_{0i}t} H_{gi}^{\alpha_{1i}} H_{si}^{\alpha_{2i}} \exp\left(\sum_j \beta_{ji} Z_{ji} + \mu_i\right)$$

where y_i is gross annual income and t is the individual's age. The number of years since the individual completed his education is H_g and it serves as a measure of general work experience; H_s is professional experience specific to his occupation; Z is a vector of dummy variables for the following characteristics:

1) level of education (Ph.D., M.A., B.A.);
2) type of employment (private industry, government, educational institution);
3) type of work activity (research, management, teaching, production and sales, consulting);
4) location of work (northeast, north-central, south, west);
5) employment status (fully or partially employed, full-time or part-time student);
6) sex;
7) citizenship and place of birth (U.S. or other).

The variable μ is an error term which describes the effect of variables such as ability, which affect earning, but which are not included in the data.

The estimate is based on observations in a single cross section, and may be interpreted as the wage structure which is expected under *static* conditions. We shall later superimpose on each wage function

TABLE 4—REGRESSION COEFFICIENTS AND PARTIAL CORRELATION COEFFICIENTS OF THE WAGE FUNCTION

	Regression Coefficients			Partial Correlations									
Field	Age	Years Since Degree	Experience	Age & Experience	Region of Employment	Citizenship & Place of Birth	Sex	Education	Employment Status	Type of Employment	Type of work Activity	No. of Scientists with Positive Income	Multiple Correlation (corrected R^2)
Chemistry	.0028*	.058	.107	.280	.010*	.002*	.019	.161	.069	.038	.247	1100	.624
Physics	.0014*	.144	.056	.277	.004*	.005*	.062	.190	.187	.083	.106	555	.702
Earth Sciences	.0007*	.080	.189	.071	.001*	.008*	.015*	.025	.176	.029	.139	540	.662
Mathematics	−.0009*	.084	.158	.277	.007*	.000*	.068	.170	.216	.075	.136	500	.688
Biology	−.0016*	.109	.118	.241	.006*	.005*	.053	.133	.056	.059	.076	727	.553
Agriculture	.0005*	−.009*	.212	.259	.000*	.009*	.007*	.118	.046	.041	.042	296	.406
Psychology	−.0019*	.034*	.140	.127	.002*	.000*	.058*	.107	.312	.049	.070	420	.591
Social Sciences	−.0028*	.068	.192	.256	.006*	.017*	.006*	.099	.131	.043	.056*	367	.558
All Fields	−.0009*	.080	.139	.232	.003	.001*	.026	.136	.119	.067	.070	5075	.569

* Insignificant at significance level .01.

the estimated rate of growth in earnings. The particular functional form (11) was adopted for pragmatic reasons. It is linear in the logs, and can be estimated by simple least square methods. Yet it allows both for decreasing marginal productivities, and for the single peaked earning profile which is usually assumed.

In Table 4 we present, for each field, the regression coefficients of the three continuous variables (t, H_g, H_s) and the coefficients of partial correlation for the other seven characteristics. Age, general work experience, and specific occupational experience are, of course, closely correlated, and one can hardly expect to obtain significant estimates for their coefficients. Nevertheless, a consistent pattern may be noted. The effect of age, as such, on the individual's earning power is insignificant, while both types of experience have a significant effect. The relatively small decay rates indicate that the peak in the earning power of the typical scientist is attained very late, if at all. Comparing the regression coefficients of general an specific experience, we note that specifi experience seems to be more productive Both types of experience exhibit decreasing marginal productivities.

Taken together, age and experience ar the most "important" explanatory variables.[11] The level of education and th type of employment and work activit are also of major importance in explainin the earning power of the typical scientist On the other hand, factors such as sex place of birth, citizenship, and locatio of work turn out to be insignfiicant.

As these results indicate, the variable which were emphasized in the wag hypothesis and the theoretical analysi which followed it are, in fact, empiricall relevant. It should be pointed out, how

---

[11] We use the partial correlations as measures o importance. The partial correlation with respect to set of explanatory variables is the increase in the regression sum of squares due to their inclusion, divide by the residual sum of squares before they were introduced. See A. Goldberger, pp. 176–78 and 197–200.

TABLE 5—REGRESSION COEFFICIENTS OF THE MODIFIED WAGE FUNCTION

Field	Constant Term	Years in the Labor Force	Level of Education is M.A.	Level of Education is Ph.D.	Full-Time Student	Part-Time Student	Multiple Correlation (corrected R^2)
Chemistry	4.293	0.172	0.156	0.377	−0.186	−0.022*	.610
Physics	4.211	0.197	0.092	0.431	−0.282	0.031*	.700
Earth Sciences	4.182	0.223	0.090	0.284	−0.244	−0.043*	.599
Mathematics	4.384	0.171	0.108	0.438	−0.100*	−0.021*	.666
Biology	4.130	0.182	0.088	0.400	−0.354	−0.016*	.541
Agriculture	4.152	0.168	0.176	0.392	−0.257*	−0.065*	.353
Psychology	4.398	0.103	0.055*	0.347	−0.317	−0.053*	.574
Social Sciences	4.400	0.172	0.041*	0.377	−0.270	−0.249	.524
All Fields	4.260	0.174	0.139	0.404	−0.249	−0.018*	.552

* Insignificant at significance level .05.

ever, that a considerable portion of the variations in scientists' gross income remains unexplained. This residual would probably be lower if the data contained some operational measure of ability.[12]

The wage function (11) may be used to predict the earning power of scientists with different types of career pattern. Our purpose, however, is to predict the earning power of scientists who follow a "straight" career, that is, they invest in education first and then work at one occupation for the rest of their lives. In this case, $H_g = H_s = H$, and $t = t_0 + E + H$, where t_0 is the age of entry into graduate school. With these a priori restrictions, we derive the wage function (12), which is a modified version of equation (11).[13]

$$(12) \qquad y_i = H_i^{\alpha_i} \exp\left(\sum_j \beta_{ij} Z_{ji} + \mu_i \right)$$

Other things equal, the earning power of the individual at each point of time in his life is determined uniquely by the level of his *experience* (measured as the number of years since his highest degree was obtained). As seen from the multiple correlations (see the last column of Table 5), the fit of the modified wage function is almost as good as the original fit. This reflects the fact that the deleted variables, age and experience, are closely correlated with the remaining variable years since degree. In other words, the restriction of straight career pattern is satisfied by most scientists in the sample.

The resulting regression coefficients, which serve as a basis for the estimation of the rate of return from schooling, are presented in Table 5. The exponent of the constant term in the first column of Table 5 is an estimate of the geometric mean of the gross starting salary of nonstudents with B.A. degrees. (All other characteristics are at their mean value.) If the distribution of earnings is lognormal, then this geometric mean is also the median.[14]

The regression coefficients in column 2

---

[12] Similar models applied to cross section data attained a similar quality of fit. N. Arnold Tolles used precisely the same statistical model and almost identical data (The 1964 Register) in his analysis of the earnings of economists. He reports R^2 of about .55. Giora Hanoch uses a linear model, and the 1960 census of population data, in the analysis of the earnings of all individuals with positive earnings. He reports R^2 of .345. A linear model applied to our sample yields a similar R^2, .342.

[13] Strictly speaking, a term $e^{\alpha_{0i} H_i}$, should be also included in (12). Since the α_{0i} are all very small, I dropped this term from the wage function in order to simplify the estimation.

[14] The lognormal distribution fits quite well to the income distributions of the scientists in the sample. See Weiss (1968, ch. 4).

of Table 5 are estimates of the elasticity of wage with respect to experience in each field. They combine the effects of all types of experience, as well as age. We note that these elasticities are all positive and smaller than one. This implies that the earnings increase with experience (age), but at a decreasing rate.

The regression coefficients for M.A. and Ph.D. degrees in Table 5 measure the net proportional contributions of acquiring these degrees to the geometric mean of earnings at each level of experience (or the median, if earnings are lognormally distributed). The Ph.D. degree yields a significant increase in earning power in all fields. The contribution of the M.A. degree may be quite small, and in two fields—psychology and social sciences—we cannot reject the hypothesis that it is, in fact, zero. In any case, the net contribution of education, holding other things constant, far exceeds the gross contribution. For instance, the median income in 1966, all fields, was $13,200 for Ph.D., $10,700 for M.A. and $11,000 for B.A. The proportional gross contribution of the Ph.D. is approximately 20 percent in contrast to about 40 percent when other things are held constant. For the M.A., the numbers are −3 percent and 14 percent, respectively.

One should not ignore the possibility that these estimates of the net contribution of education are statistically biased. The omission of variables such as ability which may be positively correlated with the level of education may cause an overestimation of the net contribution of education. Furthermore, there is the problem of simultaneity: the level of education is an endogenous controlled variable. Consequently the simple least square estimates of the wage function will in general be biased. It is interesting to note, however, that under the particular multiplicative form which we adopted for the wage function there will be no bias. This follows from the convenient property of separability possessed by the wage function (12). The rates of return, and thus the optimal levels of education, for each field are independent of the level at which we hold other things constant.[15] It is built into the model, for instance, that males and females, western or eastern, able or less able individuals, all enjoy the *same* rate of return. (Their lifetime earnings would, of course, differ.)

The type of employment and work activity are also controlled endogenous variables, and the maximizing individual would choose, for each level of education, that type of employment and work activity which maximizes earnings. If, as a result of additional education, the type of employment and work activity which is most profitable changes, then holding these variables constant would produce an underestimation of the effect of education on earnings. In order to determine the effect of education on the profitability of the various types of employment and work activity, the wage function for "all fields" was reestimated separately for each of the three education groups (B.A., M.A. Ph.D.). It turns out that only one type of employment (private industry) and one type of work activity (management) are the most profitable for *all* levels of education. It follows that no bias is caused by the inclusion of the type of employment and work activity as explanatory variables

---

[15] Let

$$e^{\delta_i t} f_i(A, H_i, E_i) = e^{a_i} H_i^{\alpha_i} e^{\beta_i(E_i)+\mu_i}$$

as in (12) where

$$\beta_i(E_i) = \begin{cases} 0 & \text{if the level of education is B.A.} \\ \beta_{1i} & \text{if the level of education is M.A.} \\ \beta_{2i} & \text{if the level of education is Ph.D.} \end{cases}$$

Suppose, as we shall later assume, that $S_i = K_i e^{\delta_i t} f_i$ wher $K_i < 1$ is a proportionality factor. Then substitutin into the first-order condition (4), we see that a_i and cancel out and thus do not pose any restriction on E_i^*.

quite the contrary—their exclusion would result in a severe underestimation of the monetary contribution of education (as well as a considerable reduction in the quality of fit). For various reasons, the relatively highly educated scientists are more attracted towards the academic type of employment and the teaching work activity where salaries are 20 percent to 30 percent lower than in private industry.[16] Consequently, a failure to hold constant the type of employment and work activity results in many cases in negative estimates of the net pecuniary contribution of education.

2) It was difficult to obtain reliable data on students' earnings by field. For lack of better data, we had to estimate these earnings from our own sample. Not all graduate students are included in the National Register, and the criteria for inclusion differ from field to field. The resulting estimates which are presented in Table 6 should be read with these qualifications in mind. It is assumed that in each field the students' net earnings are *proportional* to the scientist's earning capacity out of school.[17] As seen, the ratio of student to nonstudent earnings varies from about 39 percent in chemistry up to 64 percent in mathematics. In some fields, mathematics in particular, the resulting estimates of students' earnings (see Table 6) seem to be inflated, but overall, they compare favorably with other sources. A 1965 survey, for instance, reports average annual income of graduate students of about \$5,900, and academic expenditures of approximately \$850.[18]

3) The data on the amounts of time which are required for the M.A. and Ph.D. degrees in the various fields were obtained from the National Academy of Sciences Survey of Earned Doctorates.[19] It is easier to obtain precise data on the total time interval between the acquisitions of the various degrees, but a considerable bias results from the fact that many of the students do not acquire education continuously. Therefore, the Survey of Earned Doctorates also provides estimates of the "registered time," i.e., years of school attendance from the entry to graduate school to the acquisition of the degree.[20]

The average registered time spans from B.A. to M.A. and from M.A. to Ph.D. as reported by scientists who received their degrees in the years 1964–66 are presented in Table 7. When compared with the

---

[16] See Weiss (1968, ch. 4).

[17] The proportionality factor k_i (see column 3 of Table 6) is derived from the formula

$$k_i = x1_i e^{\beta_{18_i}} + x2_i e^{\beta_{19_i}},$$

where $x1$ and $x2$ (columns 1 and 2 of Table 6) are the proportions of full-time and part-time students with positive income from all students (including those without income). The regression coefficient β_{18_i} and β_{11_i} (see columns 5 and 6 in Table 5) measure the net proportional decrease in the earning power during the schooling period. These coefficients seem to be suspiciously small (in physics, for instance, part-time students earn more than fully employed scientists) but it should be remembered that they refer only to students who reported positive income.

[18] See "Academic and Financial Status of Graduate Students, Spring 1965." The average income was calculated from Table 1 in this study, as a weighted average of the earnings of full-time students, \$4685, and part-time students, \$6745. The percentage of full-time students in the sample is 44 percent. The average of the bracket '10,000 and more' was put at 12,000. These income figures include income of spouse which accounted for 17 percent of the full-time students' and 5 percent of the part-time students' income. The average academic expenses were claculated from Table 8, in the same study. The average for the income bracket '1700 and more' was put at 2000. The academic expenses for full-time students was \$903 in the social and behavioral sciences, and \$835 in the natural sciences.

[19] For a description of this source see: "Doctorate Recipients from U.S. Universities 1958–1966," the preface and pp. 64–80.

[20] Estimates of registered time requirements are available only for Ph.D.'s. To obtain an estimate of the registered time needed to acquire an M.A. degree, we assumed the following equality:

$$\frac{\text{Registered time B.A.} - \text{M.A.}}{\text{Registered time B.A.} - \text{Ph.D.}} = \frac{\text{Total time B.A.} - \text{M.A.}}{\text{Total time B.A.} - \text{Ph.D.}}$$

Table 6—Estimates of Gross Annual Income of Graduate Students

Field	Proportion of Students with Positive Income		Ratio of Student to Nonstudent Earnings	Annual Gross Income of Students (in dollars)				
	Full-Time Students	Part-Time Students		First Year[a]	Second Year[b]	Third Year[b]	Fourth Year[b]	Fifth Year[b]
Chemistry	.077	.331	.388	2838	3737	4007	4210	4374
Physics	.150	.283	.405	2731	3432	3718	3934	4111
Earth Sciences	.317	.383	.615	4030	5146	5633	6006	6313
Mathematics	.186	.477	.635	5093	6388	6847	7192	7472
Biology	.177	.380	.496	3084	3821	4113	4334	4514
Agriculture	.067	.600	.608	3903	5229	5597	5875	6099
Psychology	.180	.538	.641	5213	5916	6168	6353	6501
Social Sciences	.296	.444	.572	4660	5469	5865	6162	6403
All Fields	.151	.375	.486	3442	4462	4788	5034	5233

[a] Level of education is B.A.
[b] Level of education is M.A.

norm, i.e., the prescribed time span of full-time study, in most graduate schools, these figures seem surprisingly high. This is due to the practice of the Survey of Earned Doctorates not to distinguish between part-time and full-time registered students. Clearly, the amount of time spent in work directly involved in securing the degree will be closer to the norm. As seen in Table 7 there is only a small variation in the time requirements of the different fields. It is also interesting to note that, over time, these requirements remain remarkably stable.[21]

Having thus completed the estimation of lifetime earnings, student earning, and schooling periods, we are now in a position to construct the estimates of the rates of return from graduate education.

In Table 8 (columns 1 and 2) we present the estimates of the marginal and average rates of return for education in the several scientific fields.[22] We also present estimates for two special cases: in the first (columns 3 and 4), it is assumed that net earnings

---

[21] See "Doctorate Production in United States Universities 1920–1962," pp. 40–43, and also "Doctorate Recipients from United States Universities 1958–1966," Table 14.

[22] The marginal rates of return are calculated from the first-order condition (4′) when f_i is specified in (12). The marginal productivity of education is estimated by dividing the difference in earnings at two successive levels of education by the difference in their time requirements. Thus ρ, the marginal rate of return at the Ph.D. level of education in each field, is given by

$$(4'')\qquad kE_2^{\alpha}e^{\beta_1} - \frac{T'^{\alpha}e^{\beta_2}}{(1+\rho)^{T'}} + \left(\frac{e^{\beta_2}-e^{\beta_1}}{E_2-E_1} - \rho e^{\beta_2}\right)\sum_{J=1}^{T'}\frac{J^{\alpha}}{(1+J)^{\rho}} = 0$$

where k is the proportion of students' earnings from nonstudent earnings (Table 6, col. 3); E_1 is the number of years required for the M.A. degree (Table 7 col. 1); E_2 is the number of years required for the Ph.D degree (Table 7, col. 2); α is the elasticity of wage with respect to experience (Table 5, col. 2); β_1 is the ne contribution of the M.A. degree (Table 5, col. 3); β_2 is the net contribution of the Ph.D. degree (Table 5 col. 4); T' is equal to 45, the assumed maximal number of years for participation in the labor force, minus the rounded value of E_2. The marginal rate of return at the M.A. level of education is obtained from (4″) substituting E_2 by E_1, E_1 by 0, β_2 by β_1, and β_1 by 0. The average rate of return is obtained from the zero profits condition when profits are defined as in (9), f_i is specified as in (12). The average rate of return at the Ph.D level in each field is then given by

$$(9')\qquad ke^{\beta_1}\left(\sum_{J=1}^{M}\frac{J^{\alpha}}{(1+\gamma)^{J}} + \frac{M^{\alpha}}{10}\sum_{J=10M+1}^{10E_2}\frac{1}{(1+\gamma/10)^{J}}\right) + \sum_{J=1}^{T'}e^{\beta_2}\frac{J^{\alpha}}{(1+\gamma)^{J+E_2}} - \sum_{J=1}^{45}\frac{J^{\alpha}}{(1+\gamma)^{J}} = 0$$

where M is the integer part of E_2. The average rate o return for the M.A. is obtained from (9′) substituting E_2 by E_1, β_2 by β_1, and β_1 by 0.

TABLE 7—LENGTH OF SCHOOLING PERIOD
(Registered Time)

Field	(1) B.A.–M.A.	(2) B.A.–Ph.D.
Chemistry	1.7	5.2
Physics	1.9	6.1
Earth Sciences	1.7	5.4
Mathematics	1.7	5.4
Biology	1.8	5.9
Agriculture	1.6	5.5
Psychology	2.0	6.3
Social Sciences	1.7	5.9
All Fields	1.8	5.7

Source: National Academy of Sciences.

during the schooling period are zero and in the second (columns 5 and 6), it is assumed that all students complete their education in the "normal," i.e., prescribed, time of one year for the M.A. and 2.5 additional years for the Ph.D.[23] The main purpose of these additional calculations is to illustrate the considerable sensitivity of the results to alternative specification of the cost structure. This is particularly important, since the precision and reliability of the cost data is far from perfect. Each of these estimates should be interpreted as the rates of return on investment in education, which are expected under static conditions by an individual who is typical (representative) to a given field, with the average level of ability of the scientists in that field.

Let us examine first the rates of return for education, ignoring students' earnings, and assuming static expectations. These rates of return are fairly low and for the M.A. program in the social sciences they even become negative. The estimate of an average rate of return for the Ph.D. program of 6.67 percent corresponds closely to the findings of Hanoch who utilized the 1960 census, and ignoring student earnings, estimated the average rate of return from "higher education" at 7 percent.

To ignore students' earnings completely is, in the case of *graduate* students, unwarranted, especially since the schooling period used in our calculations reflects the considerable proportion of part-time students.[24] The inclusion of students' earnings approximately doubles the rates of return. The average rate of return from the Ph.D. program in all fields is in this case about 12 percent. This estimate is closer to that reported by Gary Becker for college graduates. The reduction of the length of schooling period to normal time has roughly the same effects on the rates of return as the inclusion of students' earnings. Also, for a wide range of interest rates lifetime earnings are roughly equal, whether school is completed in the prescribed time (full-time study) or in average time (with part-time work). It seems that there is little incentive for the average student to accelerate his schooling up to normal time.

So far we have ignored the expectation of growth in earnings with the passage of time. The rates of growth in median earnings during the period 1957–68 (given in the last column of Table 8), may serve as a rough estimate of these expectations. If the form of the earnings distributions remains stable, and if the same rates of growth are expected to continue in the future, all the rates of return in each field should be increased by the corresponding rate of growth.[25] The nominal average

---

[23] B. Berelson, pp. 156–76, estimates that in the physical sciences 3.5 is the median "... number of years in full-time equivalent spent in work directly involved in securing the degree (including work on the dissertation)."

[24] Among the students enrolled for master's and higher degrees in the fall of 1965, about 65 percent of the first year students, 38 percent of the intermediate students, and 41 percent of the terminal students were part-time students. See "Enrollment for Master's and Higher Degrees, Fall 1965."

[25] It is implicitly assumed that in each field the expected rate of growth in earnings is uniform for all age groups and for all levels of education. In fact, median earnings of Ph.D.'s grew faster than those of B.A.'s

(over)

TABLE 8—MARGINAL AND AVERAGE RATES OF RETURN

Field	Degree	Registered Time		Registered Time No Student Earnings		Normal Time No Student Earnings		Annual Rate of Growth in Median Earnings 1957–68[a]
		Marginal (1)	Average (2)	Marginal (3)	Average (4)	Marginal (5)	Average (6)	
Chemistry	M.A.	11.34	13.27	8.11	9.29	14.39	16.88	4.1
	Ph.D.	7.24	11.04	4.50	6.97	7.38	11.17	
Physics	M.A.	5.37	5.80	3.23	3.66	8.45	9.38	4.7
	Ph.D.	8.93	10.38	5.93	6.67	11.30	12.98	
Earth Sciences	M.A.	7.71	8.16	3.81	4.08	8.23	9.12	4.5
	Ph.D.	8.13	10.05	3.10	4.31	6.28	7.92	
Mathematics	M.A.	10.66	11.95	5.18	5.72	10.03	11.25	5.0
	Ph.D.	13.69	18.47	6.94	8.05	11.04	13.22	
Biology	M.A.	6.31	6.73	3.35	3.66	8.05	8.91	5.9
	Ph.D.	9.59	11.37	5.55	6.29	10.48	11.94	
Agriculture	M.A.	17.44	22.41	8.94	11.45	16.10	19.22	4.7
	Ph.D.	8.70	17.04	3.46	6.70	7.20	11.68	
Psychology	M.A.	4.36	5.18	0.25	0.52	4.48	4.86	6.1
	Ph.D.	10.99	12.43	4.85	4.78	9.84	10.19	
Social Sciences	M.A.	1.69	0.36	−0.76	−2.66	2.46	2.71	5.9[b]
	Ph.D.	11.02	11.66	5.90	5.70	11.24	11.17	
All Fields	M.A.	10.59	12.16	6.63	7.54	12.89	14.86	4.7
	Ph.D.	8.70	12.33	4.83	6.67	8.94	12.08	

[a] *Source:* "American Science Manpower," 1956–58, 1964, and 1968.
[b] This figure is for the period 1964–1968.

rate of return from the Ph.D. program would be, in this case, approximately 17 percent.

Let us now consider the differences in the average rates of return to education across fields. When students' earnings and the duration of the schooling period is the same in all fields (columns 6 and 5 of Table 8), we find that the rates of return from the Ph.D. program are fairly uniform (around 12 percent). The main exception is the field of earth sciences where the rate of return is distinctly lower (about 8 percent). On the other hand, there is considerable variation in the rates of return from the M.A. program. These vary from 2.7 percent in the social sciences to 19.2 percent in agriculture. Naturally, when the variation in students' earnings and the schooling period is introduced (columns 1 and 2 of Table 8), there is an increase in the variation in the rates of return. In this case the Ph.D. program offers the highest rates of return (about 18 percent) in mathematics and agriculture.

The expected decreasing marginal rates of return to education are found in only

---

between 1957 and 1968. (The annual average rates of growth in "all fields" are 5.2 and 4.1, respectively.) The correction for growth suggested in the text is therefore likely to be biased downwards. We made no correction for general price increase; the resulting rates of return should therefore be interpreted as nominal rates (comparable to the nominal rate of interest).

two fields—chemistry, agriculture—and in "all fields." In physics, earth sciences, mathematics, biology, psychology, and social sciences, there is increasing marginal productivity of education, in the sense that the marginal return to the Ph.D. exceeds that to the M.A.[26] These findings may reflect the statistical bias in our estimates of the net contributions of the Ph.D. and M.A. degrees, and of the time required to obtain them. It is possible, for instance, that holders of a M.A. degree are on the average of lower ability than those with B.A. or Ph.D. degrees. Nevertheless, one cannot reject the possibility that the marginal productivity of education is, in fact, increasing. This contrasts with our finding that the marginal productivity of experience is decreasing.[27]

There may be a certain threshold effect. That is, employers who rely on formal education as a screening method to predict ability and success on the job may consider the difference between M.A. and B.A. too small to be significant.

The Optimal Amount of Education

Given the market rate of interest, one can utilize the marginal and average rates of return to estimate the optimal amount of education for each field. In the fields where the marginal rate of return is decreasing, and is below the average rate of return, the optimal amount of education is given by the condition of equality between the marginal rate of return and the rate of interest. (We assume that, in the relevant range, the marginal rates of return curves are straight lines.) In the fields where the marginal rates of return are increasing, there will be a corner solution at the B.A. or the Ph.D. level of education, depending on whether the rate of interest is above or below the average rate of return at the Ph.D. level of education.[28]

In Table 9 we present for several rates of interest, in the presumably appropriate range (5 to 15 percent), the optimal level of education E^*, and the present value of lifetime earnings associated with it. The figures which are not in parentheses correspond to static expectations, while those in parentheses take into account the expected rates of growth in earnings.

At a rate of interest of 10 percent, the optimal level of education in most fields is at its highest feasible level (the Ph.D. degree); at a rate of 12.5 percent it is optimal in most fields not to invest in graduate education at all. This knife-edge property is, of course, the consequence of increasing returns to education. The same phenomena exist when growth is expected, except that the critical rate of interest would be between 15 and 17.5 percent.

There is a considerable variation in the lifetime earnings which can be enjoyed in the different fields, when education is carried to its optimal level. Mathematics and the social sciences are the most profitable (under static expectations), while biology and agriculture are the least profitable occupations. The ranking of the

---

[26] Under our assumptions, the marginal productivity of education is increasing whenever $(e^{\beta_2}-e^{\beta_1})/(E_2-E_1) > (e^{\beta_1}-1)/E_1$. It follows that for the marginal productivity to increase, it is sufficient that $(\beta_2-\beta_1)/\beta_1 > (E_2 - E_1)/E_1$. Reexamining the data in Tables 5 and 7, it is immediately seen that this condition is satisfied in all the fields where increasing marginal and average rates of return are found.

[27] Several other studies also report increasing return at the graduate level of education, e.g., W. Lee Hansen (1963). For an international comparison of the pattern of the rates of return from education see Hansen (1970). The most striking finding, however, is that by Orley Ashenfelter and Joseph Mooney who found the contribution of M.A. to be very small even when ability is held constant.

[28] The Ph.D. is not necessarily the upper bound on the amount of schooling. In the natural sciences postdoctoral programs are available and common. Their existence may indirectly reflect the increasing returns to education. Unfortunately, there is no information in the National Register on postdoctoral schooling.

Table 9—The Optimal Amount of Education and Present Value of Lifetime Earnings

Field	$r=0.05$		$r=0.075$		$r=0.10$		$r=0.125$		$r=0.15$	
	E^*	$V(E)^*$	E^*	$V(E)^*$	E^*	$V(E)^*$	E^*	$V(E)^*$	E^*	$V(E)^*$
Chemistry	5.2	$231,350	5.0	$149,857	2.8	$105,995	0.7	$ 83,032	0	$ 65,221
	(5.2)	(565,296)	(5.2)	(318,296)	(5.2)	(196,338)	(4.2)	(131,204)	(2.1)	(95,944)
Physics	6.1	228,430	6.1	146,004	6.1	100,056	0	77,258	0	62,826
	(6.1)	(669,668)	(6.1)	(362,691)	(6.1)	(215,591)	(6.1)	(139,057)	(6.1)	(96,007)
Earth Sciences	5.4	221,388	5.4	144,610	5.4	101,627	0	78,951	0	63,964
	(5.4)	(600,510)	(5.4)	(331,509)	(5.4)	(201,973)	(5.4)	(134,041)	(0.0)	(96,180)
Mathematics	5.4	277,030	5.4	182,335	5.4	128,873	5.4	96,359	5.4	75,272
	(5.4)	(840,972)	(5.4)	(459,144)	(5.4)	(277,030)	(5.4)	(182,335)	(5.4)	(128,873)
Biology	5.9	202,553	5.9	131,441	5.9	91,470	0	69,212	0	56,405
	(5.9)	(799,918)	(5.9)	(418,362)	(5.9)	(241,425)	(5.9)	(152,236)	(5.9)	(103,480)
Agriculture	5.5	207,745	5.5	137,754	5.0	98,441	3.9	74,127	2.7	59,439
	(5.5)	(571,276)	(5.5)	(319,722)	(5.5)	(196,925)	(5.5)	(131,783)	(4.9)	(94,279)
Psychology	6.3	215,683	6.3	143,911	6.3	102,973	0	77,939	0	64,231
	(6.3)	(852,709)	(6.3)	(450,894)	(6.3)	(264,584)	(6.3)	(170,317)	(6.3)	(118,410)
Social Sciences	5.9	256,659	5.9	167,824	5.9	117,714	0	88,941	0	72,586
	(5.9)	(999,106)	(5.9)	(525,306)	(5.9)	(305,124)	(5.9)	(193,833)	(5.9)	(132,791)
All Fields	5.7	229,214	5.7	149,561	2.7	103,257	0	77,619	0	63,328
	(5.7)	(647,292)	(5.7)	(357,460)	(5.7)	(216,867)	(5.7)	(142,795)	(2.2)	(99,140)

Note: E^* is the optimal number of years of education.
$V(E)^*$ is the present value of lifetime income which is associated with the optimal level of education (in dollars).
The numbers in parentheses are corrected for differences in the expected rate of growth in earnings.

different fields is essentially independent of the rate of interest, even in the range of 10–12.5 percent where the amount of education is sensitive to changes in the rate of interest.

Given the rate of interest, the introduction of expectations of growth in earnings affects both the amount of education and the ranking of the different occupations. Psychology and biology improve their ranking, while chemistry becomes less attractive. The figures in Table 9 clearly illustrate the importance of the growth factor. Earth sciences and psychology, for instance, are almost equivalent under static conditions ($r=0.10$) but differ by more than 30 percent when lifetime earnings are adjusted for growth.

The ranking of the different fields is not independent of our (tentative) estimates of students' earnings. When students' earnings are excluded, physics becomes somewhat more attractive and chemistry and psychology less so.

III.

That monetary motives guide occupational and educational choices is a fundamental hypothesis in the theory of human capital. For the purpose of empirical application this "monetary" hypothesis must be modified in two ways. First is the *ceteris paribus* clause: nonpecuniary advantages or disadvantages must be held constant. Second, the hypothesis can be related only to the observed behavior of groups or equivalently "representative" individuals. Even this rather humble hypothesis is by no means self-evident. Contrary to many economic problems which occur repeatedly during the individual's life, the choice of occupation is essentially unique and irreversible. The individual is typically uncertain of his preferences and abilities. He must rely on the experience and judgment of others. Frequently he is tempted to defer his decision and avoid the responsibility of conscious choice completely.

It is against this background that we raise the question whether the actual occupational and educational choices of scientists can be "explained" in terms of our estimates of the rates of return and lifetime earnings in different occupations.

The Observed Choices of New Entrants

We start with some observations on the occupational and educational choices of new entrants. New entrants are defined as scientists with less than five years experience. Their occupational choices are indicated by their rate of entry into the different fields. (The rate of entry is in each field, the proportion of scientists who are new entrants.) The educational choices of new entrants are reflected in their distribution by the level of the highest degree achieved. In Table 10 we present the rate of entry and the proportion of new entrants who possess the Ph.D. degree in each field, for the years 1964, 1966, and 1968. It is assumed that the 1966 pattern of rates of return and earnings differentials had been sufficiently stable to affect the choices which are revealed in these years.

In a state of equilibrium with fixed occupational distribution, all occupations would be equally attractive and grow at the same rate. The rates of entry would be equal across fields. As seen in Table 10 there is a tendency between 1964 and 1968 towards an equalization of the rates of entry in the different fields. Nevertheless, there are several marked changes in the distribution of scientists among fields. The question is whether the gaining fields are those which offer higher lifetime earnings.

TABLE 10—RATES OF ENTRY AND THE PROPORTIONS OF NEW ENTRANTS WITH PH.D.

(In Percent)

	Rates of Entry			New Entrants With Ph.D.		
Field	1964	1966	1968	1964	1966	1968
Chemistry	19.9	22.8	18.6	30.2	29.8	28.0
Physics	29.1	32.4	24.8	23.2	26.1	35.4
Earth Sciences	16.2	17.6	16.3	17.6	17.3	17.9
Mathematics	20.1	20.4	18.2	21.7	22.3	30.3
Biology	17.4	16.6	18.0	49.4	57.0	54.0
Agriculture	17.2	13.6	16.5	14.5	15.5	11.6
Psychology	16.0	15.8	16.5	53.0	57.2	52.3
Social Sciences	14.4	15.4	19.3	36.9	37.9	30.0
All Fields	19.3	20.6	18.7	31.1	32.8	32.4

Source: "American Science Manpower," 1964, 1966, 1968.

The lowest rate of entry is observed in agriculture which is also among the least profitable occupations (see Table 9). But low rate of entry is also observed in the social sciences which is among the most profitable fields. Similarly, physics, where the highest rate of entry is observed, offers below average lifetime earnings. In these last two fields, however, there is a possibility of lagged response indicated by the tendency of the rates of entry to decrease in physics and to increase in the social sciences.

The absence of positive correlation between rates of entry and (estimated) lifetime earnings is, of course, not necessarily inconsistent with the monetary hypothesis. First, there is an identification problem: the proportion of new entrants is fairly high and may in turn affect the observed pattern of earnings differentials. Then there is the eternal difficulty of hold-

TABLE 11—MEASURES OF ACADEMIC ABILITY

Field	Percentage of Students with Grade Point Average[a] A,A−	B,B+	B−,C	Average Score in G.R.E. Aptitude Examinations[b] Verbal	Quantitative
Chemistry				551	640
Physics	22	40	38	589	695
Earth Sciences				537	599
Mathematics	25	43	33	559	668
Biology	18	41	41	521	535
Agriculture	18	44	38	421	505
Psychology	17	45	37	565	529
Social Sciences	19	40	41	540	538
All Fields	17[c]	41[c]	42[c]	543	613

[a] "The Academic and Financial Status of Graduate Students, Spring 1965."
[b] Harvey and Marco, Table 1.
[c] Includes Humanities.

ing "other things" constant. In particular, there are two factors to be taken into account—differences in ability and educational background, and nonpecuniary advantages.

It is very difficult to obtain or even define measures of the typical scientist's ability. Nevertheless, one can utilize the available measures of academic aptitude and performance as an approximation. Among scientists, in particular, these narrowly defined measures are likely to be correlated with the more general concept of ability which is relevant to the determination of the scientist's earning power.

In Table 11 we present the average grade of students who took the graduate record examination in 1963–64, by major undergraduate field of study. We also present the undergraduate grade point average of students who enrolled in 1965, by graduate field of study. As seen in the table, physics and mathematics draw the most able students while the less capable students concentrate in agriculture, biology, and the social sciences. Taking into account the differences in ability, it seems that the estimate of lifetime earnings in physics is biased upwards while the estimate of lifetime earnings in the social sciences is based downwards. This makes the high rate of entry into physics and the low rate of entry into social sciences somewhat puzzling from the pecuniary point of view.[29]

The estimates of the marginal rates of return functions in different fields at the same point of time could provide a test of the "monetary" hypothesis which is independent of the unknown rate of interest. If the marginal rates of return are decreasing, and there is no upper bound on the amount of investment, then the marginal rates evaluated at the chosen levels of education at each field would be approximately equal. (We may assume that the *average* individual in each field faces the same rate of interest.) The level at which these rates are equated would be an estimate of the interest rate "which is relevant to investment in human capital." Since, in most fields, the marginal rates of return are increasing and corner solutions are attained, such a test is impossible. Instead we consider the relation between the proportion of new entrants who have chosen the Ph.D. level of education and the average rate of return from the Ph.D. program. Presumably, the more profitable it is to acquire a Ph.D. degree, the higher will be the proportion of new entrants possessing it.

As seen in Table 10, the proportion of new entrants possessing a Ph.D. degree is high in psychology and biology and low in agriculture and earth sciences. The rate of return to education is indeed low in earth sciences, but in agriculture it is quite high, especially when we include students' earnings. (Note, however, that total earnings in agriculture are relatively low (Table 9).) The rates of return to education (adjusted for growth) in biology and psychology are only slightly above average (see Table 8). It should be noted, however, that different eligibility criteria are exercised by the different professional associations. Thus the high proportion of Ph.D.'s in psychology and biology may reflect the relatively more demanding requirements of the corresponding associations.[30] It is perhaps more meaningful, therefore, to examine the changes in the investment in schooling which occurred between 1964 and 1968. An interesting phenomenon is the sharp increase in the proportion of Ph.D.'s in physics and mathematics. It is possible that the increased investment in graduate education which is noted in mathematics was motivated by the relatively high rates of return in that field.

---

[29] Of course, physics may be more "interesting," "relevant," etc. There is no simple way to account for nonpecuniary advantages in a cross section analysis. Over time, one may hope that changes in tastes and nonmonetary factors are small relative to, or uncorrelated with, the changes in the monetary variables. An analysis of the trends in the labor market for scientists is contained in a detailed study by R. Freeman. Referring to the National Register data, he reports "A rough positive relationship (across-fields) between the rate of growth of salaries and the number of Ph.D.'s conferred during post war years" (ch 4).

This concludes our attempt to produce evidence which would bear upon the role of monetary motives in educational and occupational choices. This study was confined to a single cross section. It is hoped that pooling of the several cross sections of the National Register will yield stronger and more conclusive tests of the monetary hypothesis. First, such pooling may enable us to determine the stability (over time) of the wage function and to examine the changes which occurred in the rates of return to education. Second, it would permit the introduction of lags and adaptive expectations. Particularly useful would be the longitudinal file which is now available. This will permit the follow-up of actual earning profiles over a period of almost fifteen years.

---

[30] See "American Science Manpower 1966," pp. 203–04.

REFERENCES

O. Ashenfelter and J. D. Mooney, "Graduate Education, Ability, and Earnings," *Rev. Econ. Statist.*, Feb. 1968, *50*, 78–86.

G. Becker, *Human Capital, A Theoretical and Empirical Analysis with Special Reference to Education.* New York 1964.

B. Berelson, *Graduate Education in the United States,* New York 1960.

A. Goldberger, *Econometric Theory,* New York 1964.

G. Hanoch, "Personal Earnings and Investment in Schooling," doctoral dissertation, Univ. Chicago 1965.

———, "An Economic Analysis of Earning and Schooling," *J. Hum. Resources,* summer 1967, *2,* 310–29.

W. L. Hansen, "Total and Private Rates of Return to Investment in Education," *J. Polit. Econ.,* Apr. 1963, *71,* 128–40.

———, "Patterns of Rates of Return to Investment in Education: Some International Comparisons," O.E.C.D. conference on Policies for Educational Growth, Paris 1970.

P. Harvey and G. Marco, "Aptitude and Advance Test Scores of 1963–64 National Program Candidates, by Undergraduate Fields," Graduate Record Examination, spec. report no. 65–3, Educational Testing Service, Princeton, Aug. 1965.

R. B. Freeman, "The Labor Market for College Manpower," doctoral dissertation, Harvard Univ. 1967.

F. Lutz and V. Lutz, *The Theory of Investment of the Firm,* Princeton 1951.

N. Tolles et al., "The Structure of Economists' Employment and Salaries, 1964," *Amer. Econ. Rev.,* Dec. 1965, supp., *55,* 1–98.

Y. Weiss, "Allocation of Time and Occupational Choice," doctoral dissertation, Stanford Univ. 1968.

———, (1971a) "Learning by Doing and Occupational Specialization," *J. Econ. Theor.,* June 1971, *3,* 189–98.

———, (1971b) "Ability and the Investment in Schooling," *J. Econ. Lit.,* June 1971, *9,* 459–61.

National Academy of Sciences, "Doctorate Production in United States Universities, 1920–1962," publication 1142, National Council, Washington 1963.

———, "Doctorate Recipients from United States Universities, 1958–1966," publication 1489, Washington 1967.

National Center for Educational Statistics, "Enrollment for Master and Higher Degrees," in O.E.-54019, final reports fall 1960-fall 1965, Washington.

———, "The Academic and Financial Status of Graduate Students, Spring 1965," O.E.-54042, Washington 1967.

National Science Foundation, "American Science Manpower," reports of the National Register of Scientific and Technical Personnel, Washington 1964, 1966, and 1968.

Hospital Cost Inflation: A Study of Nonprofit Price Dynamics

By MARTIN S. FELDSTEIN*

The empirical analysis of economic behavior has long remained the study of households and profit-making firms.[1] The growing importance of nonprofit enterprises, both as producers of goods and services and as objects of government policy, now compels us to extend our traditional area of analysis. The current paper presents an empirically estimable model of the nonprofit hospital industry and uses it to analyze the problem of hospital cost inflation.[2]

The rapid rise of hospital costs has become a major problem of public policy. In 1970 hospital services cost the nation more than $20 billion. In the three years from mid-1967 to mid-1970, the cost per patient-day rose some 42 percent. Although this represents a significant acceleration, costs have been rising very rapidly for a much longer period: a day of hospital care cost five times as much in 1970 as it did in 1950. Although the nature and causes of this severe inflation are little understood, it has induced strong pressure for government action: national health insurance, expanded grants for hospital construction and modernization, manpower training programs, incentive reimbursement payments to encourage managerial efficiency, etc. Unfortunately, without a better understanding of the mechanism of hospital cost inflation it is not possible to anticipate the impact of such policies.

Most discussions of hospital cost inflation have focused on *how* inflation has occurred (e.g., more staff, higher wages, more equipment, etc.) rather than *why* it has.[3] In contrast, the model developed in the present paper emphasizes the basic causes of hospital cost inflation through time and of hospital cost variation among different areas. The analysis suggests that, despite the nonprofit nature of the hospital industry, hospital cost inflation can

---

* Professor of economics, Harvard University. I am grateful to Bernard Friedman for assistance with this research. The study was financed by a grant from the National Center for Health Services Research and Development, U.S. Department of Health, Education, and Welfare. I have benefited from comments by a referee and by participants in seminars at Harvard and Yale universities.

[1] Quantitative research on the public sector has generally ignored the study of behavior and focused on establishing technical relations (cost and production functions), searching for empirical regularities in local government expenditure, and evaluating alternative programs (cost-benefit analyses and mathematical programming models). Some recent attempts to model optimizing behavior by local governments are a significant exception (see E. Gramlich, J. Henderson, R. Inman).

[2] This hospital industry model is part of a more general model of the health care sector that is currently being developed; for an early discussion of this project, see Feldstein (1968).

[3] See, for example, *Report to the President on Medical Care Prices*. Most empirical analyses of hospital costs have examined why costs differ among individual hospitals at a particular time rather than why the general level of costs has been rising through time. In a somewhat different type of study Judith Lave and Lester Lave identified variables that are associated with hospitals whose costs have recently increased faster than the average: teaching hospitals, hospitals in rural areas, and hospitals that had below average costs in the beginning of their period of estimation. David Salkever's related study found significant relations between cost increases and the number of interns, the local service industry wage rate, and the proportion of patients with insurance. My forthcoming study provides a general survey of why hospital costs have been rising but does not present a formal model or econometric estimates. For other discussions of the economics of hospitals, see Herbert Klarman (1964, 1965, 1969), Joseph Newhouse, M. W. Reder, Burton Weisbrod.

be explained by a model of dynamic price adjustment to excess demand. Rapidly increasing wages and the expansion of hospital employment can be viewed primarily as the results rather than the fundamental causes of this inflation.

Section I presents the formal twelve equation model of the behavior of the hospital industry. The demand and price adjustment equations defined in this model are then specified in greater detail and estimated in Sections II, III, and IV using a pooled cross section of time-series for individual states for 1958 through 1967. Section V discusses some of the implications of these estimates and the direction of future research.

I. A Model of the Economic Behavior of Hospitals

This section presents an explicit dynamic model of the use, cost, and expansion of hospital inpatient services. It is convenient to describe the model by dividing the twelve equations into four groups: demand relations, price adjustment, components of cost, and expansion of capacity.

Demand Relations. The rate of hospital admissions (*ADM*) and the mean duration of stay per admission (*MS*) together constitute the per capita demand for hospital inpatient services. The same variables are likely to affect both *ADM* and *MS*: price, income, the availability of hospital facilities and alternative sources of care, the demographic composition of the population, and general attitudes about hospital care.[4] The rationale for each of these variables and the methods of empirical implementation will be discussed in the next section. For the purpose of the general theoretical model, it is sufficient to consider the role of the price variable.

Hospital care is often described as a necessity and considered to be very insensitive to price. However, the substantial variation among areas in the rates of hospitalization and in mean durations of stay for different diagnoses and procedures shows that most treatment cannot be regarded as a technically determined necessity (see Feldstein (1967) and National Center for Health Statistics (1970)). Although admission to a hospital for some diagnoses may be completely price inelastic, admission for other conditions and the mean stays for most case types are likely to be more price elastic.

The relevant price variable for studying demand is the cost to the patient net of insurance reimbursements[5] and relative to the price of other consumer goods and services. The problem of measuring this net relative price by a single variable will be discussed in the next section. For now, the gross price of a day of hospital care (P) will be specified explicitly while variables measuring the effect of insurance and of other consumer prices will be included implicitly among the other explanatory variables of the demand equations.

Denoting this set of explanatory variables by the vector $\underline{X}_t$, the admission and mean stay equations are:

$$ADM_t = ADM(P_t, \underline{X}_t) \tag{1}$$

$$MS_t = MS(P_t, \underline{X}_t) \tag{2}$$

The total hospital bed-days demanded (*BDD*) is defined by the identity

$$BDD_t = ADM_t \cdot MS_t \tag{3}$$

Price Adjustment. The supply of patient-days can change only very slowly. An increase in a hospital's number of beds requires a relatively large "lumpy" investment. Such expansion of supply is also hampered by the nonprofit status of hospitals and their limited access to normal capital markets.[6] Although the percentage occupancy of available beds can be varied, hospitals generally have a desired occupancy rate that reflects both the need for excess capacity to deal with emergencies and the relation between the occupancy rate, unit costs, and the quality of service; this desired occupancy rate (R) will be treated as exogenous in the current analysis. Equation (4) defines the effective or desired bed-days supplied (BDS) in the short run as the product of the number of beds ($BEDS$), the desired occupancy rate (R) and the number of days per year.

---

[4] It will be assumed in the current analysis that the patients' perception of the quality of care is independent of the actual services provided by individual institutions. Under reasonable assumptions, this will not change the qualitative conclusions of the analysis.

[5] The extent of hospital insurance is treated as predetermined in the current model. A more complete analysis currently in progress will examine the feedback from higher hospital prices to the increased use of insurance.

$$BDS_t = 365 \cdot R \cdot BEDS_t \tag{4}$$

Given $\underline{X}_t$ and $BEDS_t$, equations (1) through (4) together imply an equilibrating price ($\overline{P}_t$) at which there is no excess demand:[7]

$$ADM(\overline{P}_t, \underline{X}_t) \cdot MS(\overline{P}_t, \underline{X}_t) = 365 \cdot R \cdot BEDS_t \tag{5}$$

Hospitals respond to an increase in the equilibrating price by adjusting their actual price upwards. Unlike ordinary business firms, hospitals do not raise prices in order to increase profits. Nor is there opportunity in the market for hospital services for patients to offer higher prices when there is excess demand. The mechanism that causes hospital prices to rise when the equilibrium price rises is quite different.

Every hospital's administration is under constant pressure to make changes that will raise costs per patient day. The medical staff of the hospital wants more equipment, laboratory services, and professional staff with which to provide more sophisticated care to the hospital's patients. Patient comfort and satisfaction can be increased by enlarging the nursing staff. Other groups in the hospital bureaucracy —from the social work department to the dieticians—also want to increase the scope and quality of their services. There are in addition always demands for higher wage rates for current personnel. These pressures for higher cost per patient day are contained by the hospital's need to cover costs with revenues.[8] If costs are to increase, the hospital's price must rise as well; with a constant demand function, this would cause the occupancy of hospital beds to fall below the desired level. However, when the equilibrating price ($\overline{P}$) rises, both the price and the unit costs can be increased. In short, an increase in demand permits the hospital to yield to the internal pressure for higher price methods of care without reducing the percentage occupancy of beds. Prices rise toward the clearing price with a lag that reflects the hospital's uncertainty about $\overline{P}_t$ and its desire to moderate the speed of price increases. This price response can be approximated by the simple adjustment model:[9]

---

[6] Investment in hospital beds is discussed more fully below.

[7] If $\underline{X}_t$ includes lagged values of ADM and MS, reflecting habit persistence in the demand for hospital services, $\overline{P}_t$ could alternatively be defined as the price that implies no *long-run* excess demand if only ADM and MS changed. However, defining P_t as equilibrating *current* demand and supply seems more consistent with the adjustment mechanism described below.

[8] Philanthropic contributions and income from endowment may permit small operating deficits for some institutions while others try to earn small surpluses to be accumulated for future investment. The amounts involved are small enough to be neglected in the current analysis. Local government hospitals may, of course, incur much larger deficits and the analysis would have to be modified to describe their behavior specifically. In general, however, local governments will prevent their public hospitals from spending more per patient day than the cost in comparable voluntary hospitals in the same area.

[9] In the empirical estimation, a logarithmic proportional adjustment model is used. Note that equation (5) is the usual simplification of employing a partial equilibrium adjustment in the context of a general

$$P_t - P_{t-1} = \lambda(\overline{P}_t - P_{t-1}) \tag{6}$$

The Components of Cost. Most discussions of hospital inflation stress the impact of the increasing cost of hospital inputs: more personnel, higher wages, more spending for equipment and supplies, etc. The analysis of equations (1) through (6) shows that, if the number of hospital beds is constant, rising hospital prices can be explained without reference to the course of input prices and quantities.[10] It is now time to consider explicitly the relation between the hospital price and the components of hospital cost.

Equation (7) is the basic budget constraint of the hospital:

$$P_t = C_t - D_t \tag{7}$$

where C_t is the average cost per patient-day and D_t is the deficit (or surplus) per patient-day. The deficit permitted by income from endowment and philanthropic donations is generally both small and exogenous.[11]

The average cost per patient day can be defined in terms of the number of employees per patient day (N_t), the average wage rate (w_t), an index of the quantity of materials and supplies (J_t) and their price (π_t), and the interest and other fixed costs associated with previously acquired capital stock (k_t). The unit cost identity is:

$$C_t = N_t \cdot w_t + J_t \cdot \pi_t + k_t \tag{8}$$

I shall assume for now that hospitals treat w_t and π_t as exogenous and that they are in fact exogenous to the hospital industry as a whole.[12] Since equations (1) through (5) determine C_t while w_t, π_t, and k_t are predetermined, the hospital must choose the mix of labor and nonlabor inputs subject to the unit cost identity of equation (8). The actual process of bureaucratic choice is no doubt very complex. It can perhaps be summarized most efficiently by defining a variable that measures the "*Q*uality of care as perceived by the *H*ospital bureaucracy" (*QH*) and assuming that N_t and J_t are chosen to maximize QH_t subject to the cost constraint of equation (8) and the prevailing production function[13]

$$QH_t = \phi_t(N_t, J_t) \tag{9}$$

The first-order conditions for such a constrained optimum can be solved (if the usual conditions defined by the implicit function theorem are satisfied) for N_t and J_t:[14]

$$N_t = N(C_t, w_t, \pi_t, k_t) \tag{10a}$$

and

$$J_t = J(C_t, w_t, \pi_t, k_t) \tag{10b}$$

Expansion of Capacity. Equations (1)

---

equilibrium model. A more general adjustment assumption would relate the change in P to the disequilibrium in the stock of beds as well.

[10] In fact, input prices can have an indirect effect on the price of hospital care through their effect on the number of hospital beds. This is considered below in the discussion of investment in hospital capacity.

[11] Even if the "deficit" is actually a desired surplus for future investment, I shall assume that it is exogenous. Although this could easily be modified by relating the desired surplus to the planned investment in beds, the link is probably too tenuous to be empirically important. Paul Ginsburg's study of hospital investment supports this view.

[12] The empirical studies of this part of the model will also consider a variety of more complex assumptions. It would be relatively easy to modify the analysis to allow some hospital wages to be endogenous to the hospital industry although exogenous to each individual hospital. The implications of assuming that individual hospitals have monopsony power will also be investigated. For a more general discussion of hospital wage determination, including some aspects of the role of market imperfections and "philanthropic wage policy," see Feldstein (1971b, section 5).

[13] Since N_t and J_t are in intensive form (per patient-day), this analysis implicitly excludes the possibility of economies of scale.

[14] Equations (10a) and (10b) are alternatives since, when one optimal input level is defined, the other is given by the constraint of equation (8). Although it need not concern us here that a hospital may be unable to reach the optimal levels of N_t and J_t within a single period, empirical estimates of this part of the model would have to make allowance for this.

through (10) provide the description of how hospital use, price, and costs behave in the relatively short run. If the availability of hospital beds were exogenous, the model would be complete. However, in the long run hospitals do affect the supply of beds by their investment behavior. Although the total capital budget of a hospital is generally constrained by the availability of government grants and private philanthropy, the institution has substantial scope for choosing between investment in more beds and other capital expenditures (new equipment, etc.).[15]

What determines how much a hospital invests in expanding the number of beds? Equations (1) through (10) define the trade off that hospitals face between the number of beds ($BEDS$) and the quality of care as perceived by the hospital (QH). *Ceteris paribus* an increase in $BEDS$ must decrease QH: an increased supply of bed days lowers the equilibrating price ($\bar{P}$) and therefore the actual price and the sustainable cost per patient day. The lower cost per patient day in turn restricts per day expenditure on labor and nonlabor inputs and therefore the hospital's perceived quality of care. The hospital administration must therefore choose between more patient-days and a more sophisticated service.

Figure 1 shows the basic relations of equations (1) through (10) and the implied quantity-quality trade off.[16] The first quadrant shows the *ceteris paribus* relation between the hospital's price (P) and the number of bed-days demanded (BDD). The labeling of the horizontal axis implies the equality of BDD and bed-days supplied (BDS). The linear relation between price and cost per patient-day, shown in quadrant II, corresponds to a hospital with an outside income that permits a small perpetual deficit. Quadrant III summarizes equations (8), (9), and (10), showing the relation between the cost per patient-day and the maximum attainable level of quality as perceived by the hospital. These three quadrants define the opportunity locus between the number of bed-days of care and quality of care that is drawn in quadrant IV. To emphasize the derived nature of the Q-B curve, the construction of two of its points are explicitly shown.

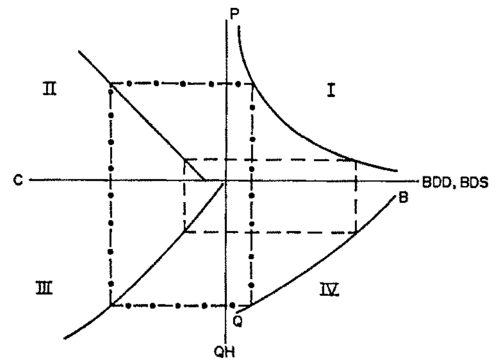


FIGURE 1

The simplest assumption about the hospital's preferences is that they can be represented by a utility function whose arguments are the number of patient bed-days (BDS) and the quality of care as perceived by the hospital:[17] $U=U(BDS, QH)$. Figure 2 combines the implied indifference curves with the opportunity locus shown in quadrant IV of Figure 1.[18] The point BDS^* indicates the number of bed-days desired by the hospital and, given the constant proportionality relation of equation (4), implies a desired stock of beds, *BEDS**. More formally, *BEDS** is defined as the stock of beds that maximizes the utility function subject to the constraints of equations (1) through (10):[19]

---

[15] Ginsburg's study showed that one aspect of the very imperfect capital market faced by nonprofit hospitals is that the ability to borrow increases with the amount of funds obtained through philanthropy and government grants. He also showed the important extent to which investment in beds and other capital spending are directly competitive.

[16] The figure represents a static model; i.e., it implicitly assumes that P adjusts to $\bar{P}$ immediately.

[17] A similar utility function has previously been incorporated into formal models by Feldstein (1967), R. Evans, and Newhouse.

[18] Although it is difficult to state the necessary conditions on equations (1)–(10) that would assure that the Q-B locus is quasi-concave from below, it is clear from the substance of the problem that an interior maximum must prevail.

$$(11)\quad BEDS^*_t = B^*(\underline{X}_t, w_t, \pi_t, k_t, D_t)$$

Equation (11) and Figure 2 indicate the indirect way that the cost of hospital inputs (w_t and π_t) can affect price. A rise in input prices will, *ceteris paribus*, lower the value of *QH* at each level of beds: the price determined by that number of beds will purchase fewer inputs and therefore yield a lower perceived quality. The *QB* curve of Figure 2 rotates downward in a counterclockwise direction. This is equivalent to changing the relative "price" of providing more bed days of care in terms of the required reduction in *QH*. Like all uncompensated price changes, its effect is indeterminate. This leads to the surprising conclusion that a rise in input prices may actually lower the price of hospital care and induce hospitals to increase the number of bed-days of care provided, if the hospital's demand for inputs has greater than unit elasticity.[20]

The actual number of beds reflects a capital stock adjustment process toward *BEDS** with the speed of adjustment depending on the availability of funds. Because this subject lies beyond the scope of the current empirical study, a detailed discussion of the availability of funds and the way that they affect investment will not be presented here. It is sufficient to write:

$$(12)\quad BEDS_t = B(BEDS^*_t, BEDS_{t-1}, \text{Availability of Funds})$$

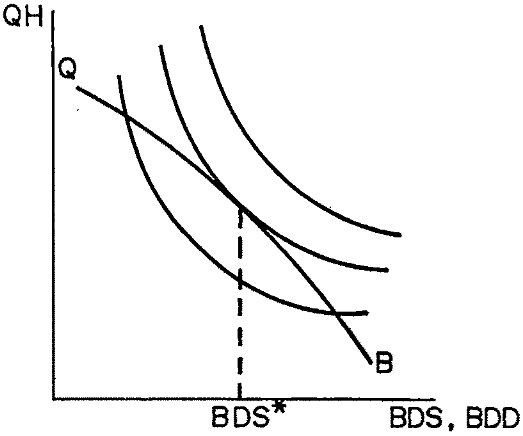


FIGURE 2

This completes the system of twelve equations that simultaneously determine the availability, use, price, and cost structure of hospital inpatient care. Although, like all models, this one involves approximations and simplifications, it is consistent with the general behavior of the hospital industry and provides a framework within which to make more specific tests and estimates.

The demand and price adjustment parts of the model have now been estimated. The specification of the demand relations (equations (1) and (2)) and the problems of definition and measurement will be discussed in the next section. Section III presents a variety of estimates of the demand relations while Section IV uses these parameters to estimate the price adjustment mechanism (equation (6)).

II. The Demand Relations: Specification and Definitions

The variables that influence the demand for admissions and mean stay were noted but not discussed in the previous section: price, income, the availability of hospital facilities and alternative sources of care, the demographic composition of the population, and general attitudes toward hospital care. Although the model of Section I described the behavior of individual patients and hospitals, data limitations require that the model be estimated for annual state aggregates. More specifically, the sample includes separate observations for each state and year in the decade 1958–67, a total of 470 observations. The precise definitions used for the empirical estimation must reflect not only the theoretical relation but also the level of aggregation, the availability of data, and the desirability of limiting the number of separate parameters to be estimated in any equation.

---

[19] This analysis assumes that the desired number of beds is independent of the interest and amortization ("user cost") per bed. Although this could easily be remedied by treating k_t as endogenous in the long-run optimization problem, it would have very little effect since the correctly defined capital cost is a very small fraction of cost per patient day (less than 10 percent). The simplifying assumption stresses that the indirect effect on quality is the important determinant of the desired bed stock.

[20] In practice, of course, a reduction in *QH* is unlikely but the input price effect may imply that *QH* and the price of hospital care grow more slowly than they otherwise would.

Price. The relative net price for each state and year was constructed using measures of the average gross price, the impact of insurance and the general level of consumer prices. Each year the American Hospital Association publishes for every state the average cost per patient day in short-term general hospitals.[21] These cost figures include interest and depreciation but exclude capital expenditures. Using average cost per patient-day as a measure of gross price is equivalent to ignoring the relatively small interstate differences in the average rate of surplus or deficit.[22]

The direct impact of insurance is to lower the *net* price paid by the patient at the time he decides how much care to consume. In practice, insurance does not merely pay a fixed proportion of hospital bills but some complex combination of proportional payments, fixed indemnities and service benefits, subject to a variety of deductions, exclusions, and ceilings. However, treating insurance as a proportional reduction in price emphasizes its primary effect and provides a useful approximation for assessing its overall impact on demand.[23] More specifically, the insurance variable (INS) is defined as the ratio of the average net price paid by consumers (NP) to the average gross price (cost) of hospitals (P).[24] Equivalently, it is the ratio of the aggregate net expenditure on short-term hospital services by consumers ($CONS$) to the aggregate total expenditure on short-term hospital services ($TEXP$) including net payments by consumers, insurance companies, and the government.[25] For state i and year t,

---

[21] The data is based on a survey of all hospitals and not a sample. The state average cost is a weighted average, weighting by the number of patient days in each hospital.

[22] The "average daily service charge" which is used as the hospital component of the consumer price index is both conceptually inferior (it omits all charges other than room and board, a large and growing part of the hospital bill) and not available for individual states.

[23] It is sometimes asserted that the effect of insurance is to lower the price elasticity of demand. There is no reason to believe that this is generally true; it depends on both the nature of the insurance and the structure of patients' preferences. If insurance paid 100 percent of hospital bills, patients would of course be completely insensitive to hospital price changes (a zero price elasticity). Other types of insurance may leave the price elasticity constant or even increase it. The approximation that insurance pays a proportion of the hospital bill illustrates this. It is perfectly possible that such insurance has no effect on the elasticity of demand; if, in the absence of insurance, the demand function has constant elasticity, this elasticity will not be changed by proportional insurance. It is also easy to see how insurance might actually increase the price elasticity of demand: at a very high price and in the absence of insurance, only medically urgent care would be purchased with the result that a small rise or fall in price would have no effect on the quantity consumed while the introduction of proportional insurance that substantially lowers the *net* price induces patients to purchase many optional items with the result that they might be quite sensitive to net price changes.

[24] The inverse of this ratio was used as the definition of insurance in the previous study of physicians' fees (Feldstein, 1970).

[25] The proportion of short-term hospital expenditure paid (or reimbursed) by insurance companies and the government rose from 68 percent in 1958 to 82 percent in 1967. If the national average total cost per patient day had remained at its 1958 value of $28, the average *net* cost to be paid by patients would have fallen from $8.96 (32 percent of $28) to $5.04 (18 percent of $28) in 1967. In fact, average total cost per patient day rose to $54, implying an 8.4 percent increase in net cost to $9.72.

$$INS_{it} = \frac{NP_{it}}{P_{it}} = \frac{CONS_{it}}{TEXP_{it}} \tag{13}$$

An explicit expression for the ratio of NP to P, that will be needed below, is:

$$\begin{aligned} INS_{it} &= \frac{NP_{it}}{P_{it}} \\ &= \frac{H_{it}UI_{it}\mu_{it} + (1 - H_{it})UN_{it}\lambda_{it}}{H_{it}UI_{it} + (1 - H_{it})UN_{it}} \end{aligned} \tag{14}$$

where H is the proportion of the population with hospital insurance, UI is the utilization rate of the insured and UN of the uninsured, μ_{it} is the proportion of their hospital bills that the insured pay themselves, and λ_{it} is the ratio of the price paid by the uninsured to the gross price.

Although the expenditure data required to calculate INS is available for the nation as a whole for every year (see L. S. Reed; D. P. Rice and B. S. Cooper), the values for individual states had to be estimated for this study. Since the proportion of the population in each state and year who have hospital insurance is available (see Health Insurance Institute, 1968),[26] the required state values of INS_{it} can be estimated by using equation (14) if the comprehensiveness of the insurance (μ_{it}) is assumed to vary with time but not among the states in each year and if the relative price for the uninsured (λ) and the relative utilization rates (UI/UN) are assumed constant. Data from a 1963 national survey of 2367 households indicated that the average values of λ and UI/UN were 0.42 and 1.2 (see R. Andersen and O. W. Anderson). The national values of μ_t for each year were obtained from equation (14) by using the national values of H_t and INS_t. Although the resulting values of INS_t are a very imperfect measure of the extent of hospital insurance, this use of partial information to represent the impact of insurance as changing the net price of care seems preferable to using only the percentage of the population with hospital insurance (H_{it}) and ignoring the national data on the changing comprehensiveness of coverage (μ_t).

The composite relative net price variable, RNP_{it}, is defined as the product of the insurance variable and the gross price divided by the national consumer price index: $RNP_{it} = INS_{it}P_{it}/CPI_t$.

The introduction of Medicaid in 1966 provided state government insurance to millions of low income families. By the end of 1966, twenty-five states had Medicaid programs; eleven more joined in 1967. The benefits, which include hospital, physician, and nursing home care, differ substantially among the states. Since many of the Medicaid eligibles previously had access to free hospital care but not to physicians and nursing homes, the expected net impact of Medicaid on hospital use by the poor is unclear. The states have not collected comparable data on the number of eligibles or the extent of use. The effect of Medicaid will therefore only be represented by a dummy variable that indicates a state's participation in Medicaid. More specifically, $MCAID_{it}$ equals 1/12 times the number of months that state i had a Medicaid program in year t. Because of this very crude measure, the estimated coefficient of the Medicaid variable should be regarded with great caution.

Income. Real per capita disposable income (INC_{it}) can be assumed to enter both the admission and mean stay demand functions in the usual way.[27] It has been suggested that low income is likely to be associated with more disease, implying a negative association between income and admissions; this will be considered when the actual estimates are presented.

---

[26] For 1966 and 1967, the number of persons with Medicare were included among the insured with prorating for the number of months of coverage in 1966. For a more specific discussion of the impact of Medicare, see Feldstein (1971a).

[27] The a priori argument that hospital care is a technical necessity and therefore completely income inelastic is wrong for the same reasons discussed in relation to price elasticity in Section I. Although the results of a 1963–64 national survey of hospital use (see National Center for Health Statistics, (1966b)) showed no income elasticity of demand for admissions

Availability. Although measures of the availability of hospital beds and of alternatives to hospital care may seem to be inappropriate in demand equations, there are good reasons for including them in the list of possible variables. Consider the impact of hospital bed availability. The fact that the number of hospital beds per 1,000 population differs substantially among areas without any sizable effect on the occupancy rate has led some observers to comment that the supply of beds "creates its own demand."[28] This important statement is unfortunately ambiguous. Does it merely imply that an increase in the number of hospital beds will, all other things being equal, lead to an increase in the quantity demanded because it depresses the price of hospital care? Or does it imply that an increase in the number of hospital beds shifts the demand schedule? Such a "pure availability effect" could operate in several ways. A relative scarcity of hospital beds may increase the waiting time for admission for elective procedures, encouraging patients to obtain ambulatory care[29] or to do without treatment. Physicians may change their own criteria for deciding when a condition "requires" hospital care and how long a stay is "appropriate" for each type of care. This may be both a reaction (conscious or unconscious) to their perception of the shortage of hospital beds and a response to pressure from the hospitals themselves. By including both the price of hospital services and the per capita supply of short-term beds (*BEDS*), it is possible to test for the presence of a pure availability effect and to assess its magnitude.

Since many medical conditions can be treated on either an inpatient or ambulatory basis, a greater availability of physicians in the area may induce relatively more ambulatory care. If the usual market mechanism were operating, a greater availability of physicians would tend to lower the price of physicians' services and therefore encourage a substitution of ambulatory for hospital care. However, a previous study (Feldstein, 1971b) indicated that the market for physicians' services does not behave as traditional theory suggests; there appears to be a persistent excess demand for physicians' services and price does not seem to vary systematically with changes in excess demand. For this reason, the availability of physicians and not their price is the appropriate variable to include in the hospital demand equations.

There is of course a different way in which local physician availability can affect hospital use: more physicians may not only encourage the substitution of ambulatory for inhospital care but may also increase the total amount of care, including inhospital care. More specifically, surgeons and other hospital-oriented doctors may induce increased demand for hospital care

---

or patient days, this evidence made no allowance for the association between income and other variables affecting hospital use. A similar 1962 national survey (National Center for Health Statistics (1966a)) showed that higher income families paid a higher price for hospital care, implying that if demand is price elastic it is also income elastic. For a further discussion of this survey data and somewhat different interpretation, see Feldstein (1971b).

[28] M. Roemer and M. Shain were probably the first to point to this. For a review of this discussion, see Klarman (1965, 1969). Although the approximate equality of "effective supply" (about 80 percent of possible total bed days) and bed use could in principle reflect a response of bed supply to the pattern of demand, there is substantial evidence from national experience in Britain (Feldstein, 1967) and local area changes in the United States (Roemer) to show that this is not so. See also the discussion of the relation between bed availability and utilization by Medicare patients in Feldstein (1971a).

[29] The term "ambulatory" includes everything but inpatient care, i.e., care in outpatient departments, doctors' offices, and patients' homes.

while general practitioners reduce that demand. For this reason, both the doctor-population ratio (*DOCS*) and the proportion of all doctors who are general practitioners (*GP*) are included in the regression.[30]

Population density (*DENS*) is another aspect of the availability of care. It is much easier to use ambulatory care in dense areas than in rural areas where patients live at substantial distances from hospitals and doctors. Increased density should therefore decrease the admission rate. Its effect on mean stay is unclear: patients can be discharged sooner in dense areas because ambulatory post-hospital care is more easily provided in dense areas but the lower admission rates may imply more serious cases requiring longer stays.[31]

Demographic Structure. The pattern of disease and the use of hospital services vary substantially by age and sex. Simple measures of the demographic structure like the proportion over sixty-five and the proportion of males are too gross and ignore important interactions. The approach used here is to adjust each admission rate (ADM_{it}) by dividing it by a demographic index based on national admission rates in sixteen age-sex-color classes,[32] and the state's demographic structure. The demographically adjusted admission rate is denoted $ADMD_{it}$. A similar adjustment is applied to the mean stay variable using a demographic index based on national mean stays; the new variable is MSD_{it}.[33]

Time. A time trend (*TIME*) was included in both demand equations to represent technical progress and changing popular attitudes about hospital care. The actual improvements in medical science and the increasing general faith in scientific medicine should both contribute to increased admissions. The effect on mean stay is less certain; one form of technical progress has been the introduction of new methods of care that reduce the length of stay for many diagnoses.

Lagged Dependent Variables. The nature of hospital use is such that individual patients cannot develop any habitual behavior with respect to admissions and mean stays. However, to the extent that patients' behavior is influenced by community norms or by their physicians, customary standards may retard adjustment to changes in the exogenous variables that influence demand. The inclusion of a lagged dependent variable in each equation provides a proportional adjustment model for testing the importance of habit and estimating the speed with which demand responds to changes in its determinants.

III. Estimates of the Demand Equations

This section presents estimates of several variants of the admission and mean stay equations.[34] These estimates serve three distinct purposes. First, they establish the importance of the traditional role for price as a determinant of the demand for hospital care. Second, they examine the effect on demand of the availability of physicians and hospitals beds. Third, they provide parameter values for use in the next section to test and estimate the price adjustment mechanism.

---

[30] Since these ratios change extremely slowly, they were treated as a constant for each state. The data for 1963 is derived from the American Medical Association.

[31] Density may also be related to different patterns of disease incidence and therefore different admission and mean stay rates.

[32] The variation by color is related less to disease incidence and more to attitude and discriminatory availability.

[33] More specifically, the sixteen national admission rates are based on a 1963–64 survey by the National Center for Health Statistics (1966b). The demographic index for each state is obtained by weighting that state's age-sex-color composition in 1960 (*U.S. Census, of Population, 1960*) by the relative national admission rates. Although the relative use rates reflect any associations between demography and availability, price, etc., the sampling and reweighting procedures makes this unlikely to be a serious source of error. Although it would be desirable to allow each state's demographic index to vary through time, the data is only available for census years. However, the effect of this added detail is unlikely to be substantial; for the nation as a whole, the demographic index changed less than one percent between 1950 and 1968.

[34] For other estimates of aggregate hospital demand relations, see Feldstein (1968); G. Rosenthal.

All of the equations are specified as linear in the logarithms of the variables (except *TIME* and the *MCAID* dummy). The equations are estimated by an instrumental variable procedure that yields consistent parameter estimates.[35] Although a separate constant term for each state is not included, the lagged dependent variables are treated as endogenous to obtain consistent estimates even if the disturbances contain a systematic "state effect" or are otherwise serially correlated.[36]

Table 1 presents the basic demand equations without measures of bed or physician availability and without lagged dependent variables. Equation (15) shows a substantial and significant price elasticity of demand for admissions, −0.63. The income elasticity is statistically significant though very small, 0.078.[37] The positive independent growth rate (3 percent) and the negative effect of density are as expected. The coefficient of the *MCAID* dummy indicates that a state that had participated for an entire year in 1967 reduced its admission demand by about 8 percent of what it would otherwise have been. The corresponding mean stay parameters in equation (16) also show a significant and substantial price elasticity (−0.49) and a much larger income elasticity (0.465). The independent growth rate and the effect of density are both very small, reflecting the countervailing factors noted in Section II. The *MCAID* coefficient is insignificant and very small. Since the total demand for bed-days (*BDD*) is the product of the admission and mean stay demands, each total demand elasticity is the sum of the corresponding *ADM* and *MS* elasticities. The total price elasticity is therefore −1.12 and the total income elasticity is 0.54.

Because the relative net price variable is defined as the product of the relative price (*RP*, the gross price divided by the consumer price index) and the insurance variable (*INS*), the specification of equations (15) and (16) is equivalent to constraining the elasticities with respect to *RP* and *INS* to be the same. The validity of this constraint depends on the assumption that treating insurance as a proportional price reduction is in fact a good approximation. Equations (17) and (18) drop the constraint and estimate separate elasticities with respect to *RP* and *INS*. In the admission equation (17), the two elasticities are quite close to each other (−0.55 and −0.68) and therefore to the original net price elasticity. The standard error of the estimate is unchanged; the other parameters are also essentially unchanged except that the income elasticity is decreased and insignificant. In the mean

---

[35] More specifically, the set of instruments includes all the exogenous variables in the two demand equations plus a number of other exogenous variables that would appear in other equations if a complete model of the health care sector were being estimated. The method is thus similar in spirit to the technique suggested by F. Fisher, but without the formal ordering and test procedure, which would, in any case, be impossible since a full model of the health sector has not been specified.

[36] Even if the *sample* disturbances differ systematically among the states, the omission of separate constant terms does not bias the regression coefficients if the "state effects" are of the transitory type studied by Marc Nerlove (see also P. Balestra and Nerlove) or, more generally, if they are uncorrelated with the explanatory variables. Although a lagged dependent variable would be correlated with the disturbance if there is such a state effect, instrumental variable estimates are consistent. Balestra and Nerlove and Nerlove show that including separate constant terms may yield inferior estimates.

[37] This very low value may reflect both the true income elasticity and a negative correlation between illness and income. Higher income individuals may also consume higher quality and higher priced hospital services, a relation that is not considered in the current model.

TABLE 1—BASIC DEMAND EQUATIONS

		Explanatory Variables							Regression Statistics	
Equation	Dependent Variable	*RNP*	*RP*	*INS*	*INC*	*TIME*	*DENS*	*MCAID*	*RSS*	*SER*
(15)	*ADMD*	−0.626			0.075	0.031	−0.071	−0.080	6.054	0.1142
		(0.048)			(0.030)	(0.003)	(0.005)	(0.024)		
(16)	*MSD*	−0.494			0.464	0.005	0.010	0.008	3.324	0.0846
		(0.035)			(0.022)	(0.002)	(0.003)	(0.018)		
(17)	*ADMD*		−0.550	−0.685	0.015	0.028	−0.075	−0.090	6.033	0.1142
			(0.106)	(0.088)	(0.081)	(0.004)	(0.007)	(0.027)		
(18)	*MSD*		−0.386	−0.580	0.378	0.000	0.004	0.006	3.474	0.0866
			(0.080)	(0.067)	(0.062)	(0.003)	(0.005)	(0.020)		

Note: Standard errors are shown in parentheses. All variables except *TIME* and *MCAID* are logarithms. The constant term for each equation is not shown.

Observations are for individual states for 1958 through 1967.

Estimation is by instrumental variables. The squared multiple correlation is therefore not shown. The final columns show the residual sum of squares and the standard error of the regression.

Explanatory variables are: *RNP* relative net price; *RP* relative price; *INS* insurance; *INC* per capita disposable income; *DENS* population density; *MCAID* medicaid dummy.

stay equation (18), the elasticities differ somewhat more (−0.39 and −0.58); the standard error of the estimate is slightly increased by dropping the constraint and the other parameters are essentially unchanged.

The equations in Table 2 assess the effect on demand of the availability of physicians and hospital beds. Equation (19) adds the doctor-population ratio (*DOCS*) and the ratio of general practitioners to total physicians (*GP*) to the variables in the previous admission equation. Since the income elasticity is extremely small (0.01) and insignificant, the equation is reestimated without it as equation (20). The coefficients of the two physician variables reflect the opposite impacts of specialists and general practitioners: an increase in the doctor-population ratio causes a small increase in admissions while an increase in the relative number of general practitioners reduces admissions.[38] The price elasticity (−0.74) is slightly greater than in equation (15); the other coefficients are unchanged. Equation (21) adds the two physician variables to the basic mean stay relation; *DOCS* is insignificant and the equation is therefore reestimated without it as equation (22). The zero elasticity with respect to *DOCS* and the negative elasticity with respect to *GP* implies that the elasticities with respect to the *GP*-population ratio and the specialist-population ratio are of equal absolute value (0.102 in 1963) and opposite sign.[39] An increase in the number of *GP*'s reduces hospital stays

---

[38] It is instructive to reinterpret the two physician coefficients of equation (20) to estimate the separate effects of general practitioners and other doctors. Let G be the ratio of general practitioners to the population and S the ratio of other doctors to the population Denoting the impact of the nonphysician variables by Z, equation (20) may be written:

$$ADMD = (G+S)^{0.056}[G/(G+S)]^{-0.124}Z$$

or

$$ADMD = (G+S)^{0.180}G^{-0.124}Z$$

The elasticities of $ADMD$ with respect to G and S are therefore $0.180[G/(G+S)]-0.124$ and $0.180[S/(G+S)]$. In 1963, the national value of $G/(G+S)$ was 0.27. Calculated at this value, the elasticity of admissions with respect to G was −0.075 and the elasticity with respect to S was +0.131.

[39] This follows from the derivation of the separate elasticities in the previous footnote.

while an increase in the number of specialists induces longer stays. The other coefficients are essentially unchanged from equation (16); the price elasticity is somewhat increased and the very small elasticity with respect to density has become insignificant and changed sign.

Equation (23) adds the per capita bed availability variable (*BEDS*) to equation (20); equation (24) reestimates this relation without the insignificant *DOCS* variable. The estimated admission elasticity with respect to *BEDS* (0.25) is highly significant, indicating the existence of a "pure availability effect." With a constant price level, an increase in the per capita bed supply induces additional demand for hospital admissions. All of the other coefficients are somewhat smaller than they were when the *BEDS* variable was not included. Similar results are obtained with the mean stay equation. Equation (25) adds *BEDS* and equation (26) then omits the insignificant *TIME* variable. The "pure availability effect" elasticity of mean stay with respect to *BEDS* is 0.28. The price elasticity, which had been −0.59 in equation (22), is reduced to −0.24 but remains highly significant.

Considering equations (24) and (26) together shows that the elasticity of total demand for bed days with respect to bed availability is 0.53. Since an increase in bed availability has almost no effect on percentage utilization,[40] it is a reasonable first approximation to say that half of an increased supply is absorbed by a pure availability effect (without a price decrease) and the other half by a fall in price.[41]

TABLE 2—EFFECT ON DEMAND OF PHYSICIAN AND HOSPITAL BED AVAILABILITY

Equation	Dependent Variable	Explanatory Variables											
		RNP	*RP*	*INS*	*INC*	*TIME*	*DENS*	*MCAID*	*DOCS*	*GP*	*BEDS*	*RSS*	*SER*
(19)	*ADMD*	−0.738 (0.057)			0.010 (0.036)	0.037 (0.003)	−0.091 (0.006)	−0.099 (0.026)	0.053 (0.029)	−0.124 (0.041)		7.004	0.1231
(20)	*ADMD*	−0.735 (0.056)				0.038 (0.003)	−0.090 (0.006)	−0.098 (0.026)	0.056 (0.026)	−0.124 (0.041)		6.973	0.1227
(21)	*MSD*	−0.593 (0.041)			0.418 (0.026)	0.010 (0.002)	−0.007 (0.005)	−0.007 (0.019)	0.025 (0.021)	−0.123 (0.029)		3.613	0.0884
(22)	*MSD*	−0.592 (0.041)			0.431 (0.024)	0.009 (0.002)	−0.006 (0.005)	−0.005 (0.019)		−0.140 (0.026)		3.622	0.0884
(23)	*ADMD*	−0.419 (0.68)				0.024 (0.003)	−0.064 (0.006)	−0.076 (0.019)	−0.011 (0.022)	−0.107 (0.029)	0.255 (0.044)	3.567	0.0879
(24)	*ADMD*	−0.435 (0.061)				0.025 (0.003)	−0.066 (0.006)	−0.078 (0.019)		−0.101 (0.027)	0.243 (0.038)	3.691	0.0893
(25)	*MSD*	−0.146 (0.072)			0.265 (0.030)	−0.0037 (0.0025)	0.029 (0.006)	0.031 (0.016)		−0.092 (0.021)	0.330 (0.047)	2.206	0.0691
(26)	*MSD*	−0.236 (0.041)			0.288 (0.026)		0.022 (0.004)	0.017 (0.013)		−0.108 (0.019)	0.277 (0.032)	2.200	0.0689
(27)	*ADMD*		−0.455 (0.051)	−0.043 (0.078)		0.031 (0.002)	−0.041 (0.006)	−0.026 (0.018)		−0.166 (0.025)	0.414 (0.041)	2.577	0.0747
(28)	*MSD*		−0.253 (0.041)	−0.127 (0.071)	0.327 (0.033)		0.028 (0.005)	0.037 (0.017)		−0.114 (0.019)	0.311 (0.036)	2.131	0.0679

See note to Table 1.

Additional explanatory variables include: *DOCS* doctor-population ratio; *GP* ratio of general practitioners to total physicians; *BEDS* per capita bed availability.

---

[40] The simple elasticity of total bed-day demand (*BDD*) with respect to bed availability (*BEDS*) is 0.92.

[41] More beds in an area may also attract more specialist physicians which would also increase demand.

Finally, equations (27) and (28) repeat equations (24) and (26) with separate price elasticities for relative price (*RP*) and insurance (*INS*). In both, the elasticity with respect to *INS* is less than the elasticity with respect to *RP*. In the admission equation (27), the elasticity with respect to *INS* has become insignificant and the elasticities with respect to *GP* and *BEDS* have increased; the *RP* elasticity (-0.46) is essentially unchanged from equation (22). The reason for this difference between the *RP* and *INS* admission elasticities is not clear. It may be a statistical artifact, reflecting the approximations in defining *INS* or its collinearity with *GP* and *BEDS*. It may however reflect a greater concern in admission decisions with the gross price than with the net price, perhaps because patients are uncertain of their insurance coverage until after they are hospitalized or even because physicians influence admissions more than mean stay and are more concerned with gross resource use than with the net cost to patients. Only further studies of a more disaggregated kind will clarify this. In the mean stay equation (28), the elasticity with respect to *INS* (-0.13) is statistically significant and the other coefficients are essentially unchanged from equation (26).

The basic specifications of equations (24) and (26) have been reestimated with lagged dependent variables to assess the speed with which demand adjusts to changes in the explanatory variables.[42] The results are presented in Table 3. The coefficient of lagged admissions in equation (29) indicates that approximately half of the full long-run response of admission demand occurs in the first year and nearly 90 percent by the end of three years. The implied long-run elasticities, i.e., the coefficients divided by $1-0.472=0.528$, are very similar to the estimates of equation (24); these are reported in square brackets below the standard errors. The coefficient of the lagged mean stay in equation (30) shows a slower speed of adjustment, approximately 40 percent in the first year and 80 percent in three years. The long-run elasticities show the same general results as equation (26) but a somewhat greater effect of bed availability and Medicaid and somewhat reduced elasticities with respect to price, income, and the *GP* variable. The lagged adjustment process may also contribute to the general impression of a very low price elasticity since the short-run price elasticity of total demand (-0.271) is less than half of the long-run elasticity (-0.553).

The important qualitative findings of this section can be summarized briefly. First, there is a substantial price elasticity of demand for hospital bed days. Second, there is also an important "pure availability effect"; i.e., an increase in the availability of beds increases quantity demanded by shifting the demand curve as well as lowering the price. Third, a high ratio of general practitioners to population lowers the demand for hospital care while a high specialist to population ratio increases demand. Finally, there is a substantial lag in the adjustment of demand to the causal variables.

IV. The Price Adjustment Mechanism

This section incorporates the parameter estimates of the previous section into the price adjustment mechanism of Section I. Rather than choose among the alternative parameter estimates on the basis of relatively small differences in the residual sums of squares, this section will investigate three different pairs of demand equations in assessing the price adjustment model: equations (24) and (26) which include physician and bed availability, equations (27) and (28) which have separate relative price (RP) and insurance (INS) elasticities, and equations (29) and (30) which add the lagged demand variables. The results for all three models provide very good explanations of the interstate and intertemporal variation in hospital prices.

---

[42] This specification imposes the assumption that the speed of response is the same for all variables. With such short time-series, it is unlikely that additional detail could be estimated by considering a more complex model.

Equation (4) of Section I described the desired short-run supply of bed days in the ith state at time t by:

$$BDS_{it} = 365 \cdot R_i \cdot BEDS_{it} \tag{31}$$

where R_i is the desired proportional occupancy rate of the available beds. The equilibrating gross price ($\overline{P}_{it}$) at which the bed day demand equals BDS_{it}, is defined implicitly by:

$$ADM(\overline{P}_{it}, \underline{X}_{it}) \cdot MS(\overline{P}_{it}, \underline{X}_{it}) = BDS_{it} \tag{32}$$

The estimates of Section III specified the admission and mean stay equations to be linear in the logarithms of the relevant variables. To implement the price adjustment model, the relative price variable (RP) must be rewritten as the ratio of the gross price to a CPI index variable ($RP_{it} = P_{it}/CPI_{it}$) and the demographically adjusted admission and mean stay variables must also be rewritten explicitly in terms of the two demographic indices: $ADMD_{it} = ADM_{it}/DEMADM_{it}$ and $MSD_{it} = MS_{it}/DEMMS_{it}$. Then, letting lower case letters denote the logarithm of the corresponding upper case variable, typical mean stay and admission equations can be written as:

$$\begin{aligned} adm_{it} = {} & demadm_{it} + \alpha_0 + \alpha_1(p_{it} - cpi_{it}) \\ & + \alpha_2 ins_{it} + \sum_{j=3}^{k} \alpha_j x_{jit} \end{aligned} \tag{33}$$

and

$$\begin{aligned} ms_{it} = {} & demms_{it} + \beta_0 + \beta_1(p_{it} - cpi_{it}) \\ & + \beta_2 ins_{it} + \sum_{j=3}^{k} \beta_j x_{jit}, \end{aligned} \tag{34}$$

where the x_{it} are the logarithms of the other variables in these demand equations. If the equations specify a composite RNP variable, $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$.

Equating the demand for bed-days ($bdd_{it} = adm_{it} + ms_{it}$) and the short-run supply at the equilibrating price ($\overline{P}_{it}$) implies:

$$\begin{aligned} & demadm_{it} + demms_{it} + \alpha_0 + \beta_0 \\ & \quad + (\alpha_1 + \beta_1)(\bar{p}_{it} - cpi_{it}) \\ & \quad + (\alpha_2 + \beta_2) ins_{it} + \sum_{j=3}^{k} (\alpha_j + \beta_j) x_{jit} \\ & = r_i + beds_{it} + log\ 365 \end{aligned} \tag{35}$$

TABLE 3—DEMAND EQUATIONS WITH LAGGED ADJUSTMENT

Equation	Dependent Variable			Explanatory Variables					*LAG DEP VAR*	Regression Statistics	
		RNP	*INC*	*TIME*	*DENS*	*MCAID*	*GP*	*BEDS*		*RSS*	*SER*
(29)	*ADMD*	−0.211 (0.054) *−0.400*		0.014 (0.003) *0.027*	−0.033 (0.007) *−0.062*	−0.058 (0.013) *−0.110*	−0.050 (0.020) *−0.095*	0.139 (0.032) *0.263*	0.472 (0.093)	1.224	0.0543
(30)	*MSD*	−0.060 (0.035) *−0.153*	0.093 (0.037) *0.237*		0.012 (0.003) *0.031*	0.034 (0.008) *0.087*	−0.027 (0.018) *−0.069*	0.132 (0.036) *0.336*	0.607 (0.114)	0.625	0.0388

Note: Standard errors are shown in parentheses. Long-run elasticities are shown immediately below the standard errors in italics. All variables except *TIME* and *MCAID* are logarithms. The constant term for each equation is not shown.

Observations are for individual states for 1959 through 1967.

Estimation is by instrumental variables. The squared multiple correlation is therefore not shown. The final columns show the residual sum of squares and the standard error of the regression.

Solving for $\bar{p}_{it}$ and substituting into the proportional adjustment mechanism

$$p_{it} - p_{i,t-1} = \lambda(\bar{p}_{it} - p_{i,t-1}) \tag{36}$$

yields:

$$\begin{aligned} p_{it} &- p_{i,t-1} \\ &= \frac{\lambda}{\alpha_1 + \beta_1}[r_i + \mathit{log}\ 365 - \alpha_0 - \beta_0] \\ &+ \frac{\lambda}{\alpha_1 + \beta_1} \\ &\cdot [\mathit{beds}_{it} - \mathit{demadm}_{it} - \mathit{demms}_{it}] \\ &+ \lambda(\mathit{cpi}_{it} - p_{i,t-1}) - \frac{\lambda(\alpha_2 + \beta_2)}{\alpha_1 + \beta_1}\mathit{ins}_{it} \\ &- \lambda \sum_{j=3}^{k} \frac{\alpha_j + \beta_j}{\alpha_1 + \beta_1} x_{jit} \end{aligned} \tag{37}$$

Although equation (37) could be estimated directly, the substantial multicollinearity would make the directly estimated parameters quite unreliable. Moreover, it is unnecessary to do so since all of the parameters except λ and r_i have already been obtained by estimating the demand equations. The estimation of λ and r_i can be simply achieved by using the previously estimated parameters to define the composite variable

$$\begin{aligned} Z_{it} = \frac{1}{\alpha_1 + \beta_1}\Bigg[&\mathit{beds}_{it} - \mathit{demadm}_{it} \\ &- \mathit{demms}_{it} - (\alpha_2 + \beta_2)\mathit{ins}_i \\ &- \sum_{j=3}^{k}(\alpha_j + \beta_j)x_{jit}\Bigg] \\ &+ \mathit{cpi}_t - p_{i,t-1} \end{aligned} \tag{38}$$

which reduces the estimation equation to

$$\begin{aligned} p_{it} - p_{i,t-1} = \frac{\lambda}{\alpha_1 + \beta_1}&[r_i + \mathit{log}\ 365 \\ &- \alpha_0 - \beta_0] + \lambda Z_{it} \end{aligned} \tag{39}$$

Equation (39) has been estimated by the same consistent instrumental variable procedure used for the demand equations. Since the r_i are presumed to differ, a separate constant term was estimated for each state. When Z_{it} is defined by the parameters of the demand equations (24) and (26), the estimated speed of adjustment (λ) is 0.250 with a standard error of 0.039: the price change in any year is approximately one-fourth of the difference between the equilibrating ($\bar{P}_{it}$) and the previous actual price ($P_{i,t-1}$). The values of $\bar{P}_{i,t}$ for each state and year were then calculated from equation (35) using the individual constant terms of equation (39) and the estimate of λ to calculate each r_i. Although the estimated speed of adjustment implies that the equilibrating price will generally not equal the actual price, the values of $\bar{P}_{it}$ provide a very good explanation of the overall variation in actual price among states and years. More specifically, the simple correlation between P_{it} and $\bar{P}_{it}$ is 0.988. Since $\bar{P}_{it}$ is a function of variables like insurance, income, and the availability of beds but not of the current or lagged values of P_{it}, this high correlation is evidence of the model's substantial explanatory power.

Very similar results were obtained when the parameters of equation (27) and (28) were used. The estimated adjustment speed is 0.280 with standard error 0.045. The simple correlation between P_{it} and $\bar{P}_{it}$ is also 0.988.

The parameters of the lagged adjustment demand equations (29) and (30) imply a much smaller short-run price elasticity of demand and therefore a much greater rise in the equilibrating price in response to a change in any of the other variables that affect demand. To be consistent with the actual price rise, the response speed in this equation is lower than in the previous equations (0.089 with a standard error of 0.017); the correlation is again very high (0.987). This model of slow adjustment to the price that equates short-run demand and supply therefore has the same empirical implications as the previous models of more rapid adjustment to the price that equates long-run demand and supply.

As the model of Section I showed, the demand and price adjustment equations would provide a complete explanation of hospital cost inflation if the supply of beds were fixed. During the decade 1958–67, the per capita increase in beds was quite small (15 percent) in comparison to the increase in price (91 percent). This suggests that, as a first approximation, the demand and price adjustment analysis alone can provide useful estimates of the impact on hospital prices of rising income, more complete insurance, etc. Such estimates must be interpreted with care: they represent the partial effect on price, assuming that the availability of bed days did not respond to shifts in demand. Similarly, the estimated effect of the actual increase in bed supply ignores the role of feedbacks through the price mechanism. A more complete analysis and a reduced form calculation cannot be done until the full model has been estimated.

Table 4 presents the current analysis of the contribution that each of the explanatory variables in the price adjustment equation made to the change in the equilibrating price for the nation as a whole. More specifically, the first row of the table shows the percentage change from 1958 to 1967 for the nation as a whole in each of the variables that influences the equilibrating price. Equation (37) was used to calculate the elasticity[43] of $\overline{P}$ with respect to each of these variables for the three sets of demand equations; these elasticities appear in the second, fourth, and sixth rows of the table. The rows marked "contribution" show the percentage increase in the equilibrium price implied by the elasticity and the percentage change of the corresponding variable. The sum of the contributions in any row is the total percentage change in $\overline{P}$ according to that model. Thus equations (24) and (26) imply that $\overline{P}$ rose 135 percent from 1958 to 1967. More than half of this increase is attributed to increased insurance coverage.[44] Since the actual price (i.e., average cost per patient day) only rose 91 percent over this period, the estimate implies that, even if there had been no further changes in the explana-

TABLE 4—CONTRIBUTIONS TO HOSPITAL COST INFLATION, 1958–67

Percentage Change* 1958–67	*CPI* 15.5	*INC* 31.1	*TIME* [9]	*DENS* 13.6	*INS* −44.5	*MCAID* [0.766]	*GP* −11.6	*BEDS* 13.8
Equations								
(24) Elasticity	1.000	0.429	0.037	−0.066	1.000	−0.091	−0.311	−0.715
and								
(26) Contribution	15.5	12.3	35.8	−0.8	80.2	−6.7	3.9	−8.9
(27) Elasticity	1.000	0.462	0.044	−0.018	0.241	0.016	−0.395	−0.388
and								
(28) Contribution	15.5	13.3	48.3	−0.2	15.3	1.2	5.0	−4.9
(29) Elasticity	1.000	0.429	0.049	−0.056	1.000	−0.042	−0.297	−0.725
and								
(30) Contribution	15.5	12.3	55.1	−0.7	80.2	−3.1	3.7	−0.0

* For *TIME* and *MCAID*, these are absolute changes. Since *INS* is defined as the ratio of net price to gross price, a fall in *INS* reflects an increase in the extent of insurance coverage.

---

[43] For *TIME* and *MCAID* the coefficients are not elasticities.

[44] Note that since *INS* is defined as the ratio of net price to gross price, an increase in insurance coverage decreases *INS*.

tory variables, prices would have continued to rise substantially after 1967. The model of equations (27) and (28) offers a more optimistic prospect. The total increase in $\overline{P}$ was only 93.5 percent. This reflects the much smaller calculated impact of insurance. Finally, the results for equations (29) and (30) are quite similar to (24) and (26) with a total rise in $\overline{P}$ of 153 percent.

V. Summary and Conclusions

Hospital cost inflation has become a significant national problem. Price increases averaging 12 percent a year from mid-1967 to mid-1970 have brought pressures for radical changes in the organization and financing of our health care system. Economic advice about the problem has been hampered by the lack of previous empirical research and, more generally, of an analytic framework for describing the behavior of the hospital industry. This in turn reflects a general lack of attention to the problems of the nonprofit sector.

The current paper attempts to contribute to our general understanding of this important nonprofit industry and to provide specific information about the determinants of hospital cost inflation. Section I presented a theoretical model of the use, pricing, and expansion of hospital services. The demand and price adjustment equations were estimated using a mixed cross section of time-series sample of ten annual observations for each state.

The estimated demand equations showed substantial price elasticities for both per capita admission and mean stay per case and an important income elasticity for mean stay. A greater availability of general practitioners appears to reduce the demand for both admissions and for longer stay while a greater availability of other doctors has the opposite effect. The availability of beds was shown to increase admission and mean stay by a "pure availability effect" as well as through lower prices. The relatively slow response speeds in the demand and price adjustment equations suggest that rather substantial price increases may continue for some time, even if there is no additional stimulus to demand.

The model as a whole and the estimated equations in particular provide an alternative to the customary explanation of hospital cost inflation. Increases in the components of cost are seen as primarily the result and not the cause of higher prices. The source of inflation is the pressure of the rising demand induced by increases in insurance coverage, personal incomes, the availability of hospital oriented specialists, etc.

As insurance companies and the government pay a higher proportion of hospital bills, patients face a lower net price; from 1958 to 1967, this proportion rose from 68.1 to 82.3 percent. The estimates presented in this paper suggest that this growth of insurance may account for as much as an 80 percent rise in the national equilibrating price. Moreover, the growth of insurance differs from the other causes of inflation in an important way. Insurance is bought to avoid the risk of unexpected expenditure but, because it provides a reduction in price at the time that care is purchased, it has the concomitant effect of artificially increasing the demand for care and its price.[45] Because hospital insurance is already so comprehensive, a further growth of our current type of insurance would not only lead to substantial price increases but might so limit price sensitivity that the system becomes unstable and requires direct controls.[46]

[45] For a very clear statement of this problem, sometimes pejoratively labeled "moral hazard," see Mark Pauly.

[46] This study has of course assumed that the price *elasticity* has not been affected by increased insurance coverage. A separate study of this issue would undoubtedly be worthwhile.

The opposite impacts of general practitioners and specialists on the demand for hospital care also has important implications. The estimates suggest that an increase in the number of general practitioners would induce a very large saving in hospital resources, on the order of $39,000 a year per general practitioner.[47] This estimate, if substantiated by more detailed studies, provides substantial support for a policy to increase the number of general practitioners or other providers of primary care. Conversely, the estimate of the increased use of hospital services induced by an additional specialist (again approximately $39,000 a year),[48] if verified by further studies, should be taken into account in any decisions to increase the number of hospital-oriented physicians.

The research described in this paper is currently being extended in several directions. The two remaining sections of the hospital behavior model (the demand for current inputs and the expansion of capacity) will be estimated. The problem of quality change and its impact on consumer demand will be examined explicitly.[49] The growth of insurance and the behavior of physicians will also be made endogenous. If these related studies are successful, the research as a whole will provide the basis for a more complete behavioral model of the health care sector.

REFERENCES

R. Anderson and O. W. Anderson, *A Decade of Health Services,* Chicago 1967.

P. Balestra and M. Nerlove, "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica,* July 1966, *34,* 585–612.

R. G. Evans, "Efficiency Incentives in Hospital Reimbursement," unpublished doctoral dissertation, Harvard Univ. 1970.

M. S. Feldstein, *Economic Analysis for Health Service Efficiency: Econometric Studies of the British National Health Service,* Amsterdam 1967.

———, "The Use of an Econometric Model for Health Sector Planning," in *Federal Programs for the Development of Human Resources:* A Compendium of Papers submitted to the Subcommittee on Economic Progress of the Joint Economic Committee, Washington 1968.

———, "The Rising Price of Physicians' Services," *Rev. Econ. Statist.,* May 1970, *52,* 121–33.

———, (1971a) "An Econometric Model of the Medicare System," *Quart. J. Econ.,* Feb. 1971, *85,* 1–20.

———, (1971b) *The Rising Cost of Hospital Care,* forthcoming 1971.

F. M. Fisher, "Dynamic Structure and Estimation in Economy-Wide Econometric Models," in J. S. Duesenberry et al., eds., *The Brookings Quarterly Econometric Model of the United States,* Chicago 1965.

P. Ginsburg, "Capital in Non-Profit Hospitals," unpublished doctoral dissertation, Harvard Univ. 1970.

E. M. Gramlich, "State and Local Governments and Their Budget Constraint," *Int. Econ. Rev.,* June 1969, *10,* 163–82.

J. M. Henderson, "Local Government Expenditures: A Social Welfare Analysis," *Rev. Econ. Statist.,* May 1968, *50,* 156–63.

---

[47] This estimate is based on the following calculation. The estimated elasticity of bed-day use with respect to the ratio of general practitioners to total doctors is approximately −0.2. Footnote 38 shows that this implies that the elasticity of bed-day use with respect to the *GP* to population ratio is −0.2 times the ratio of "other doctors" to "total doctors." In 1967, this ratio was 0.78, implying an elasticity with respect to the *GP* to population ratio of −0.156. Thus a 10 percent increase in the *GP* to population ratio would reduce bed-days used by 1.56 percent. With ten *GP*'s per 28,000 population in 1967 and 32,000 hospital days for the same population, an increase of one general practitioner would reduce bed-day use by 505 days. Since the average cost per day in 1970 is some $77, the saving in hospital costs per additional *GP* is approximately $39,000.

[48] This follows from the fact that an increase or decrease in the total doctor population ratio has no effect as long as the mix of doctors is unchanged.

[49] See Feldstein (1971b, section 4) for a theoretical model of endogenous quality change in which patients' preferences determine the prevailing technology.

R. Inman, "Three Essays on Fiscal Federalism," unpublished doctoral dissertation, Harvard Univ. 1971.

H. E. Klarman, "The Increased Cost of Hospital Care," in S. J. Axelrod, ed., *The Economics of Health and Medical Care,* Ann Arbor 1964, 227–254.

———, *The Economics of Health,* New York 1965.

———, "Approaches to Moderating the Increases in Medical Care Costs," *Med. Care,* May-June 1969, *7,* 175–90.

J. R. Lave and L. B. Lave, "Hospital Cost Functions," *Amer. Econ. Rev.*, June 1970, *60,* 379–95.

M. Nerlove, "Experimental Evidence on the Estimation of Dynamic Economic Relations from a Time-Series of Cross-Sections," *Econ. Stud. Quart.*, Dec. 1967, *18,* 42–74.

J. P. Newhouse, "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *Amer. Econ. Rev.*, Mar. 1970, *60,* 64–74.

M. Pauly, "The Economics of Moral Hazard," *Amer. Econ. Rev.*, June 1968, *58,* 531–37.

M. W. Reder, "Some Problems in the Economics of Hospitals," *Amer. Econ. Rev. Proc.*, May 1965, *55,* 472–80.

L. S. Reed, "Private Health Insurance, 1968: Enrollment, Coverage, and Financial Experience," *Soc. Sec. Bull.*, Dec. 1969, *32,* 19–35.

D. P. Rice and B. S. Cooper, "National Health Expenditures, 1929–1968," *Soc. Sec. Bull.*, Jan. 1970, *33,* 3–20

M. I. Roemer, "Bed Supply and Hospital Utilization: A Natural Experiment," *Hospitals,* 1961, *35,* 36–42.

——— and M. Shain, *Hospital Utilization under Insurance,* Chicago 1959.

G. D. Rosenthal, *The Demand for General Hospital Facilities,* Chicago 1964.

D. Salkever, "Studies in the Economics of Hospital Costs," unpublished doctoral dissertation, Harvard Univ., 1970.

B. A. Weisbrod, "Some Problems of Pricing and Resource Allocation in a Non-Profit Industry—the Hospitals," *J. Bus., Univ. Chicago,* Jan. 1965, *38,* 18–20.

American Hospital Association, *Hospitals, Journal of the Am. Hosp. Assoc., Guide Issue, Part II,* 1950–1969.

American Medical Association, *Selected Characteristics of the Physician Population, 1963–1967,* Chicago 1968.

Health Insurance Institute, *Source Book of Health Insurance Data,* New York 1959–1968.

National Center for Health Statistics, (1966a), *Personal Health Expenses: Per Capita Annual Expenses, U.S.: July-Dec. 1962,* series 10, no. 27, Washington 1966.

———, (1966b), *Hospital Discharges and Length of Stay: Short-Stay Hospitals, U.S., July 1963–June 1964,* series 10, no. 30, Washington 1966.

———, *Inpatient Utilization of Short-Stay Hospitals by Diagnosis, U.S., 1965,* series 13, no. 6, Washington 1970.

U.S. Bureau of the Census, *U.S. Census of Population: 1960,* Vol. 1 Characteristics of the Population, Part I, Washington 1964.

U.S. Dept. of Health, Education and Welfare, *Report to the President on Medical Care Prices,* Washington 1967.

Congestion Tolls Under Monopoly

By Noel M. Edelson*

One of the few unequivocal propositions in economic theory is that, for goods involving no externalities, a nondiscriminating monopolist will restrict output below the Pareto optimal level. James M. Buchanan (1969) has recently shown that a monopolist's output may be either larger or smaller than the Pareto optimal level if his production process involves external costs. The indeterminacy arises because a monopolist equates marginal revenue to private marginal cost, whereas the Pareto optimal rate of output occurs where price equals social marginal cost (private marginal cost plus external marginal cost). If a regulatory agency were to impose a tax schedule such that at every rate of output marginal tax payments equalled marginal external costs, the monopolist's private marginal cost curve would coincide with the social marginal cost curve. By equating marginal revenue to his new marginal cost curve the monopolist would produce less than the Pareto optimal rate of output.[1]

The purpose of this note is to see whether Buchanan's conclusion about the effect of production externalities also applies in the case of externalities in consumption. The problem we analyze is that of congestion, which is itself a type of collective good insofar as the delay imposed by crowding is "consumed" equally by all people using the crowded facility.

Consider the following model of commuting via alternate transportation modes, which is based upon the problem first considered by A. C. Pigou and Frank Knight. Let there be a train which takes suburbanites to the city in one hour for a "net" fare of one dollar. We assume that this travel time is independent of the number of commuters, and that the one dollar net fare equals the marginal cost of carrying an additional commuter less the direct costs (gasoline, parking) of commuting by car. Instead of riding the train suburbanites can travel by car along a highway which is subject to congestion delays. Commuting by car takes $t(N)$ hours, where N is the number of cars (with one commuter per car) entering the highway per unit time.

$$(1) \quad t(N) > 0, \quad t'(N) \geqq 0, \quad t''(N) \geqq 0$$

Assume further that each commuter has an opportunity cost of commuting time equal to V dollars per hour; V is constant during an individual's trip and is the same for train and car travel. Allowing V to vary with the length of the trip would add nothing to a model in which V differs as between individuals. The constancy of V implies that, for commuter j, the value of time spent in commuting by car is $V(j)t(N)$ when commuters are entering the highway at rate N, whereas it is always $V(j)$ if j travels by train.

In Section II we derive the well-known result that the highway toll charged by a monopolist is Pareto optimal if $V(j)$ is independent of j. We relax this assumption in Section III to show that Pareto optimality and profit maximization generally imply different tolls if commuters have different values of time, and that a monopolist's toll may admit a traffic flow larger than the Pareto optimal rate. The concluding section considers a model in which both travel modes are subject to congestion delays. There we demonstrate that if $V(j)=V$ a monopolist will restrict traffic flow below the Pareto optimal rate, but if commuters have different values of time the opposite can occur.

---

* Assistant professor of economics, University of Pennsylvania.

[1] The monopoly rate of output is always below the Pareto optimal level if the production process confers external benefits. If external benefits were internalized, either by a governmental subsidy or by some market-based compensation scheme, the monopoly rate of output would still be too small.

Our model is clearly unrealistic in that the arrival rate of commuters is taken to be deterministic, but similar results would obtain in a stochastic version if individual utility functions were linear in income. This is so because we are concerned with equilibrium allocations of commuters to highway and train travel, i.e., situations in which no commuter has an incentive to switch to another mode of transportation because it is cheaper.

I. Congestion Tolls with Homogeneous Values of Time

Suppose that commuters leave for the city at rate L per unit time; N of these drive and $L-N$ take the train. Cars will crowd onto a toll-free highway until the one dollar net train fare plus the monetary value of time spent in train travel just equals the monetary value of time spent in commuting by car. The equilibrium value of N, denoted N_c, is given by

$$Vt(N_c) = 1 + V \tag{2}$$

or

$$N_c = t^{-1}\left(\frac{1+V}{V}\right) \tag{2'}$$

If $t'(\ ) > 0$, $t^{-1}(\ )$ and N_c will be unique.

The equilibrium N_c is not Pareto optimal, because a small change $dN_c < 0$ increases social welfare. The net social gain due to shifting a commuter from car to train travel is approximately[2] $-VN_c t'(N_c)dN_c$. The Pareto optimal value of N, denoted N^*, is obtained either by maximizing commuter utility, defined as the difference between costs of train and highway travel for those commuting by car, or by minimizing total travel costs. The former expression is $N[1+V-Vt(N)]$ and the latter $NVt(N)+(L-N)(1+V)$. Differentiating with respect to N we obtain

$$\begin{aligned} VN^*t'(N^*) &= 1 + V - Vt(N^*) \\ &= 1 + V[1 - t(N^*)] \end{aligned} \tag{3}$$

The left-hand side of (3) is the value to N^* drivers of reducing travel time by $t'(N^*)$ hours; the right-hand side is the additional cost imposed on the commuter shifted from highway to rail travel. Pareto optimality requires that these be equal, and hence that $N^* < N_c$. Comparing the toll-free and Pareto optimal equilibria, there would be a net social gain of $N^*V[t(N_c)-t(N^*)]$ if a costless way could be found to shift $N_c - N^*$ drivers to the train.

Now suppose that a nondiscriminating monopolist is given the right to collect highway tolls. If marginal cost to the monopolist of another car using the road and toll booths is zero, and if toll booths do not delay traffic, the monopolist will set a toll that maximizes his revenue. For any toll T, the equilibrium number of cars entering the highway per unit time is given by[3]

$$T(N) = 1 + V[1 - t(N)] \tag{4}$$

The value of N which maximizes $T(N)N$ $\bar{N}$, is given by

$$\bar{N}Vt'(\bar{N}) = 1 + V[1 - t(\bar{N})] = T(\bar{N}) \tag{5}$$

Expression (5) implies that $\bar{N} = N^*$, the result first proved by Knight. Unlike

---

[2] Walters provides estimates of the gain by reducing congestion on English highways, pp. 172–76, and urban roads, pp. 177–79; 203–05.

[3] We ignore the dynamic process by which this equilibrium might be reached. Conceptually we might imagine that the monopolist announces a toll, T, and a maximum number of cars to be admitted, N. If N is defined by equation (4), travel time of $t(N)$ hours is assured. Therefore, N people would be willing to pay the toll.

Buchanan's model of a production externality, the consumption externality of crowding is internalized by conferring the property right of collecting tolls. The monopolist takes account of the delays due to crowding because these reduce the toll he can collect from commuters.

It follows that monopoly profits equal our previous measure of maximum social gain when $N^* = \overline{N}$.[4] Consequently commuters have no reason to oppose the toll, while recipients of the toll revenue clearly favor it.

II. Congestion Tolls with Heterogeneous Values of Time[5]

Begin by arraying the available flow of commuters, L, in decreasing order with respect to their individual values of time. This yields a schedule $V(k)$ such that

$$\begin{aligned} V(k) &\geqq 0, \qquad 0 \leqq k \leqq L \\ V'(k) &\leqq 0, \qquad 0 \leqq k \leqq L \end{aligned} \tag{6}$$

In equilibrium, when the highway toll is T, for all N automobile commuters the following inequality must hold:

$$T + \mathrm{t}(N)V(k) \leqq 1 + V(k), \qquad k \in N;$$

whereas for all $L-N$ train commuters

$$T + \mathrm{t}(N)V(j) \geqq 1 + V(j), \qquad j \in L - N$$

The following two propositions enable us to prove that the toll-free equilibrium cannot be Pareto optimal:

PROPOSITION 1: *If $T=0$, we cannot have an equilibrium where $t(N) \leqq 1$ unless $t(L) \leqq 1$.*

PROOF:

For any commuter j we have $1+V(j) > V(j) \geqq V(j)\mathrm{t}(N)$ if $\mathrm{t}(N) \leqq 1$. Therefore, every commuter has an incentive to switch to the highway if travel time by car does not exceed one hour.

PROPOSITION 2: *If $T=0$ in equilibrium, $1+V(L-N) = V(L-N)t(N)$ and, for all commuters k using the highway, $V(k) \leqq V(L-N)$.*

PROOF:

Let commuter j be the one indifferent between car and train travel, i.e., $1+V(j) = V(j)\mathrm{t}(N)$. Now consider another commuter x with valuation $V(x) = V(j)+e > V(j)$. Since $\mathrm{t}(N) > 1$ by Proposition 1, $1+V(x) = 1+V(j)+e$ is less than $V(j)\mathrm{t}(N) + e\mathrm{t}(N)$. Therefore, commuter x prefers to commute by train rather than car. It follows that all automobile commuters have a valuation of time no greater than commuter j. But if N commuters are using the highway, $j = L-N$.

Propositions 1 and 2 imply that, in a toll-free equilibrium, travel time by highway will exceed one hour, and the highway commuter just indifferent between car and train has the highest valuation of time of his fellows. Switching commuter $L-N$ to the train yields a social gain larger than $NV(L)\ \mathrm{t}'(N)$, so this equilibrium cannot be Pareto optimal.

It is also easy to show that $T=1$ cannot be Pareto optimal. When the highway toll and train fare are equal everyone will prefer the highway if driving time is less than one hour, and the train if driving time exceeds one hour. Provided that $\mathrm{t}(L) > 1$, the equilibrium flow of cars can only be $\hat{N} = \mathrm{t}^{-1}(1)$. This, in turn, implies that $1+V(j)\mathrm{t}(\hat{N}) = 1+V(j)$ for all j, so that $\hat{N}$ car commuters can be drawn from anywhere along the schedule $V(k)$. Switching any one of these commuters yields a net social gain larger than $\hat{N}t'(\hat{N}) \min_{j\epsilon\hat{N}} V(j)$.

---

[4] If there were some fixed costs associated with constructing toll booths, the net social gain of setting a Pareto optimal toll on a previously toll-free road would be correspondingly smaller.

[5] Martin Beckmann et al. considered a special case of our model where commuters choose between roads of equal capacity. With heterogeneous values of time it is easily seen that $N^* < L/2$, but they did not investigate the relationship between Pareto optimal and profit-maximizing tolls.

The boundary value $T=1$ divides the set of possible Pareto optimal tolls into two regions: $T>1$ and $0<T<1$. Let N be the equilibrium flow of cars corresponding to every toll, T. By a proof similar to that for Proposition 2 we can show that, if $T>1$, $t(N)<1$ and $V(j)\geqq V(N)$ for all $j\epsilon N$; if $T<1$, $t(N)>1$ and $V(j)\leqq V(N)$. In other words, for tolls in excess of one dollar N is defined by $T+V(N)t(N)=1+V(N)$, and for tolls less than one dollar by $T+V(L-N)t(N)=1+V(L-N)$. In each case increasing the toll reduces the equilibrium flow of auto traffic. If $T>1$, the one who leaves with a higher toll is the driver with the lowest value of time, whereas if $T<1$, the one who leaves with the higher toll is the driver with the highest value of time.

A Pareto optimal flow of traffic is one which maximizes "commuter net utility," defined for this model as the difference between the costs of train and car travel for those using the highway. This measure is either

$$\int_0^N \{[1+V(j)] - t(N)V(j)\}dj \equiv U_1(N) \tag{7}$$

or

$$\int_{L-N}^L \{[1+V(j)] - t(N)V(j)\}dj \equiv U_2(N), \tag{8}$$

depending on whether N is defined by $T+V(N)t(N)=1+V(N)$ or $T+V(L-N)\cdot t(N)=1+V(L-N)$.[6]

Differentiating (7) and (8) with respect to N and setting the derivatives equal to zero we obtain

$$\int_0^{N^*} t'(N^*)V(j)dj = 1+V(N^*)[1-t(N^*)] \tag{9}$$

$$\int_{L-N^*}^L t'(N_*)V(j)dj = 1+V(L-N_*)[1-t(N_*)] \tag{10}$$

Equations (9) and (10) are the same as (3) when $V(j)=V$. The left-hand sides of (9) and (10) are nondecreasing in N^* and N_*, respectively, while the right-hand sides, since $t(N^*)<1$ and $t(N_*)>1$, are nonincreasing. Excluding the possibility of $dt/dN=dV/dN=0$ at equilibrium, we conclude that N^* and N_* are unique.

After substituting the maximizing values of N^* and N_* back into (7) and (8) we determine which of these expressions is larger. If $U_1(N^*)>U_2(N_*)$ the Pareto optimal toll is

$$T^* = 1+V(N^*)[1-t(N^*)] > 1, \tag{11}$$

whereas if $U_2(N_*)>U_1(N^*)$ the Pareto optimal toll is

$$T_* = 1+V(L-N_*)[1-t(N_*)] < 1 \tag{12}$$

An equilibrium at N^*, the relatively smaller flow of traffic consisting of high

---

[6] Note that maximizing net utility of highway users is equivalent to finding the least-cost allocation of commuters to travel modes. Since we know that the N highway users will be those with highest (lowest) values of time if N^* is less than (exceeds) $\hat{N}$, N^* is defined by either (7′) or (8′).

$$\min_N \int_0^N t(N)V(j)dj + \int_N^L [1+V(j)]dj \tag{7′}$$

$$\min_N \int_0^{L-N} [1+V(j)]dj + \int_{L-N}^L t(N)V(j)dj \tag{8′}$$

But because

$$\int_N^L [1+V(j)]dj = \int_0^L [1+V(j)]dj - \int_0^N [1+V(j)]dj,$$

$$\int_0^{L-N} [1+V(j)]dj = \int_0^L [1+V(j)]dj - \int_{L-N}^L [1+V(j)]dj,$$

and $\text{minimum}_N\, f(N) = \text{maximum}_N -f(N)$, we find that (7′) and (8′) are identical to (7) and (8).

time value commuters, is more likely: 1) the more inelastic is the schedule $V(k)$ at high values of time; 2) the larger is $\mathrm{t}''(N) \geqq 0$.

A Pareto optimal toll will not be accepted as readily as before, however. When individuals have equal values of time, toll revenue equals the gain in social welfare; commuting costs are unchanged from the pretoll situation, while recipients of the toll revenue experience an increase in welfare. With heterogeneous values of time all former highway commuters are worse off as compared with the toll-free case. This is because it is the low time value individuals who use the highway when $T=0$.

The toll-free equilibrium is one in which highway travel time exceeds one hour and highway traffic flow, N_c, is defined by the condition $1+V(L-N_c)=V(L-N_c)\mathrm{t}(N_c)$. If the Pareto optimal toll is less than one dollar, so T_* is defined by (12), N_c-N_* drivers will switch to the train while N_* of the lowest time value commuters remain on the highway. An individual with time value $V(L-N_c)\geqq V(L-N) > V(L-N_*)$ experiences a decrease in net utility equal to $1+V(L-N)[1-\mathrm{t}(N_c)]$ by being priced off the highway. An individual with time value $V(L-N) \leqq V(L-N_*)$ continues to drive, but experiences a decrease in net utility equal to $T_*-V(L-N)[\mathrm{t}(N_c)-\mathrm{t}(N_*)]$. Since $V'(\ ) \leqq 0$, these losses are larger the smaller the commuter's value of time. Since N_* is the Pareto optimal traffic flow, toll revenue exceeds total losses. Even so, some riders might be worse off after imposing the toll if it is infeasible to discriminate in terms of compensation offered, i.e., if all losers have to receive the same amount. The revenue T_*N_* will not be large enough to pay all N_c former highway users the maximum loss, $T_* - V(L)[\mathrm{t}(N_c) - \mathrm{t}(N_*)]$, if $V(L) < T_*(N_c - N_*)/N_c[\mathrm{t}(N_c) - \mathrm{t}(N_*)]$.

If the Pareto optimal toll exceeds one dollar, all former highway users shift to the train, and the N^* highest time value commuters shift to the highway. A former highway user with $V(L-N) \leqq V(L-N_c)$ loses all of his net commuter utility, $1+V(L-N)[1-\mathrm{t}(N_c)]$; the largest such loss is experienced by the commuter with the lowest value of time, $V(L)$. Each new highway user with $V(N) \geqq V(N^*)$ has consumer surplus equal to

$$1 + V(N)[1 - \mathrm{t}(N^*)] - T^* = [V(N) - V(N^*)][1 - \mathrm{t}(N^*)],$$

where formerly he had none. Therefore, new highway users as well as recipients of the toll revenue favor imposition of a Pareto optimal toll, whereas former highway users unanimously oppose it. As in the previous case, consumer surplus plus toll revenue is sufficient to compensate losers if it is possible to pay more to commuters with lower values of time.

If toll revenue is not sufficient to pay all losers the maximum loss, it certainly will not provide enough funds if lump sum transfers are infeasible. Given such a restriction the simplest way to compensate losers is to subsidize train fares and/or expand highway capacity. The desirability of the later option depends on the long-run marginal cost of highway versus rail transport, evaluated at the Pareto optimal commuter flows. Subsidizing train fares is an "inefficient" procedure in that it benefits train riders who have experienced no losses. Moreover, lowering the train fare by S dollars requires a corresponding decrease in the highway toll if traffic patterns are to remain Pareto optimal. The maximum subsidy, S, that can be paid out of Pareto optimal toll revenue is given by $(T-S)N=S(L-N)$, which implies that $S^*=T^*N^*/L$ and $S_*=T_*N_*/L$.

Now consider the toll a nondiscriminating monopolist will set to maximize his profits, assuming once again that he perceives the marginal cost of another commuter using the highway to be zero. The monopolist's average revenue curve, $T(N)$, is given by

$$\begin{aligned} T(N) &= 1 + V(N)[1 - \mathrm{t}(N)], \\ &\qquad 0 \leqq N \leqq \hat{N} = \mathrm{t}^{-1}(1) \\ &= 1 + V(L-N)[1 - \mathrm{t}(N)], \\ &\qquad N \geqq \hat{N} \end{aligned} \tag{13}$$

If $V(N)$ and $\mathrm{t}(N)$ are continuous there will be a cusp in the average revenue curve at $\hat{N} = \mathrm{t}^{-1}(1)$ if $\hat{N} \neq L/2$ and $V'(\hat{N}) < 0$. At the cusp point $T(N)$ is continuous but not differentiable. This is because $V(L-\hat{N}) > V(\hat{N})$ if $\hat{N} > L/2$ and $V(L-\hat{N}) < V(\hat{N})$ if $\hat{N} < L/2$. The situation with $\hat{N} > L/2$ is depicted in Figure 1.

The monopolist's marginal revenue curve is given by

$$1 + V(N)[1 - \mathrm{t}(N)] + NV'(N)[1 - \mathrm{t}(N)] - NV(N)\mathrm{t}'(N)$$

for

$$0 \leqq N \leqq \hat{N}$$

and

$$1 + V(L-N)[1 - \mathrm{t}(N)] - NV'(L-N) \cdot [1 - \mathrm{t}(N)] - NV(L-N)\mathrm{t}'(N)$$

for

$$\hat{N} \leqq N \leqq L$$

The properties of this function are elucidated in Appendix A. There we show that profit maximization is likely to imply either (14) or (15).

$$\overline{N}V(\overline{N})\mathrm{t}'(\overline{N}) - \overline{N}V'(\overline{N})[1 - \mathrm{t}(\overline{N})] = -\overline{N}T'(\overline{N}) = T(\overline{N}) > 1 \tag{14}$$

and $T'(N)N + T(N) < 0$ for $N > \hat{N}$

$$T'(N)N + T(N) > 0 \qquad \text{for } N < \hat{N} \tag{15}$$

and $\underline{N}V(L - \underline{N})\mathrm{t}'(\underline{N}) + \underline{N}V'(L - \underline{N}) \cdot [1 - \mathrm{t}(\underline{N})] = T(\underline{N}) < 1$

If (14) holds the profit maximizing toll exceeds one dollar and is given by

$$T(\overline{N}) = 1 + V(\overline{N})[1 - \mathrm{t}(\overline{N})], \tag{16}$$

whereas if (15) holds, the profit maximizing toll is given by

$$T(\underline{N}) = 1 + V(L - \underline{N})[1 - \mathrm{t}(\underline{N})] \tag{17}$$

Comparing (9) and (10) with (14) and (15) we see that maximizing net utility is equivalent to maximizing monopoly profits if, and only if, $V(k) = V$ for all k.

The Pareto optimal and monopoly outputs generally do not coincide if commuters have different values of time, and it is possible for the latter to exceed the

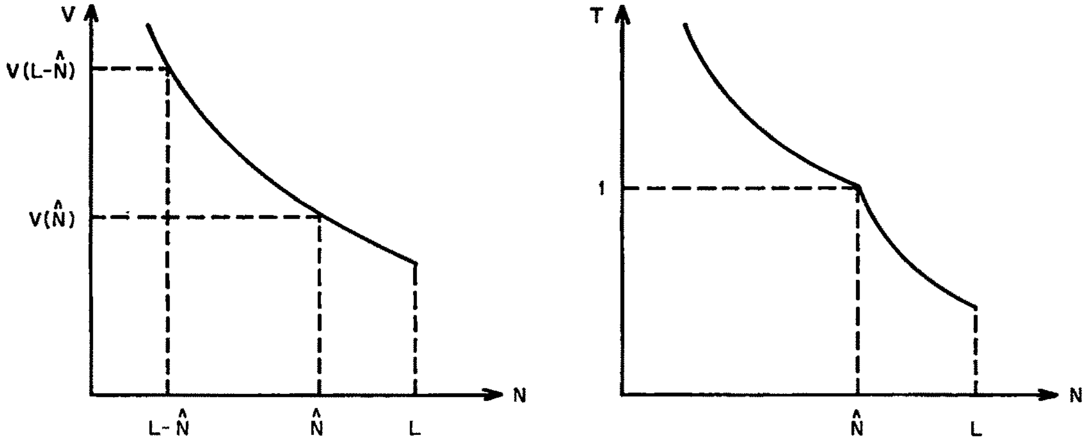


FIGURE 1

former. This is because a socially efficient allocation of commuters measures the marginal cost of an additional car by the reduction in consumer surplus of existing users; these losses will be larger the larger are drivers' values of time. A nondiscriminating monopolist ignores these differences in $V(j)$, and considers only the decrease in highway toll needed to induce an additional commuter to switch from train to car.

For example, suppose both Pareto optimality and profit maximization require a highway toll greater than one dollar, which implies that N^* and $\overline{N}$ are less than $\hat{N}$. For $\overline{N}$ to exceed N^* marginal revenue, $T(N)+NT'(N)$, must be positive at N^*:

$$T(N^*) > N^* \left| V'(N^*) \right| [1 - t(N^*)] + N^*V(N^*)t'(N^*) \tag{18}$$

Since N^* satisfies (9),

$$\begin{aligned} T(N^*) &= \int_0^{N^*} t'(N^*)V(j)dj \\ &= N^*t'(N^*)V(N^* - M), \\ &\qquad\qquad 0 < M < N^* \end{aligned} \tag{19}$$

Combining (18) and (19) we find that a necessary condition for $\overline{N} > N^*$ is that

$$\begin{aligned} N^*t'(N^*)[V(N^* - M) - V(N^*)] \\ > N^* \left| V'(N^*) \right| [1 - t(N^*)], \end{aligned} \tag{20}$$

which can be satisfied if $V(N^*-M)$ is large enough relative to $V(N^*)$.

In the Appendix we construct an example where the Pareto optimal flow of traffic exceeds that flow which maximizes profits, and another example where the reverse is true. This implies that if commuters have unequal (or nonconstant) marginal valuations of time, a monopolist may set a toll which is too large or too small. A congestion toll above the monopoly level is desirable only if the Pareto optimal flow of highway traffic is smaller than that which maximizes profits.

III. Concluding Remarks

The possibility of congestion on both travel modes transforms our model into one involving reciprocal externalities; accordingly, some of the previous results are altered. We continue to assume that the (net) direct cost of another rail passenger is one dollar, and that this equals the train fare; $t_1(N)$ and $t_2(L-N)$ denote highway and railroad travel times when N commuters drive and $L-N$ take the train.

Using the alternate formulation presented in footnote 6, our problem is to select N so as to minimize

$$\int_0^N V(j)t_1(N)dj + \int_N^L [1 + V(j)t_2(L - N)]dj \tag{21}$$

or

$$\int_{L-N}^L V(j)t_1(N)dj + \int_0^{L-N} [1 + V(j)t_2(L - N)]dj, \tag{22}$$

which yield the first-order conditions

$$\begin{aligned} &\int_0^{N^*} V(j)t_1'(N^*)dj - \int_{N^*}^L V(j)t_2'(L - N^*)dj \\ &= T(N^*) \\ &= 1 + V(N^*)[t_2(L - N^*) - t_1(N^*)] \end{aligned} \tag{23}$$

or

$$\begin{aligned} &\int_{L-N_*}^L V(j)t_1'(N_*)dj - \int_0^{L-N_*} V(j)t_2'(L - N_*)dj \\ &= T(N_*) \\ &= 1 + V(L - N_*) \\ &\quad \cdot [t_2(L - N_*) - t_1(N_*)] \end{aligned} \tag{24}$$

Equations (23) and (24) differ from (9) and (10) in that the left-hand sides represent the *net*[7] increase in the cost of travel time when a commuter is shifted from rail to highway.

The first change we note is that it is possible, though unlikely, for the toll-free equilibrium to be Pareto optimal. With $T=0$ commuters enter the highway at rate N_c, defined by

$$V(L - N_c)\mathrm{t}_1(N_c) = 1 + V(L - N_c)\mathrm{t}_2(L - N_c),$$

and this may be Pareto optimal if[8]

$$\int_{L-N_c}^{L} V(j)\mathrm{t}_1'(N_c)dj = \int_{0}^{L-N_c} V(j)\mathrm{t}_2'(L - N_c)dj$$

Another difference is that the monopoly and Pareto optimal equilibria do not coincide if commuters have the same values of time. If $V(j) = V$ for all j, the monopolist's average revenue and marginal revenue curves are $T(N) = 1 + V[\mathrm{t}_2(L-N) - \mathrm{t}_1(N)]$ and $T(N) - NV[\mathrm{t}_2'(L-N) + \mathrm{t}_1'(N)]$, respectively. The first-order conditions for Pareto optimality and profit maximization are

$$N^*V\mathrm{t}_1'(N^*) - (L - N^*)V\mathrm{t}_2'(L - N^*) = T(N^*) \tag{25}$$

and

$$\overline{N}V[\mathrm{t}_1'(\overline{N}) + \mathrm{t}_2'(L - N)] = T(\overline{N}), \tag{26}$$

which imply that $T(N^*) < T(\overline{N})$. Monopoly output is too small in this case because admitting another car requires lowering the highway toll not only due to increased highway congestion but also due to reduced travel time on the train. This time saving *reduces* the net cost of additional highway commuters from society's point of view.

With heterogeneous values of time, however, it is still possible for a monopolist to admit more than the Pareto optimal flow of highway traffic. Assuming that $\mathrm{t}_2 > \mathrm{t}_1$ in equilibrium, and rewriting the integrals in (23) by the mean value theorem, we have the following first-order conditions:

$$\begin{aligned} &N^*V(N^* - M)\mathrm{t}_1'(N^*) \\ &\quad - (L - N^*)V(N^* + K)\mathrm{t}_2'(L - N^*) \\ &= T(N^*), \qquad M > 0, \qquad K > 0; \\ &\overline{N}V(\overline{N})[\mathrm{t}_1'(\overline{N}) + \mathrm{t}_2'(L - \overline{N})] \\ &\quad - \overline{N}V'(\overline{N})[\mathrm{t}_2(L - \overline{N}) - \mathrm{t}_1(\overline{N})] \\ &= T(\overline{N}) \end{aligned} \tag{27}$$

The inequality $T(\overline{N}) < T(N^*)$ can occur if $V(N^* - M)$ is large enough and $V(N^* + K)$ small enough relative to $V(\overline{N})$.

The possibility that profit maximization can lead to excessive congestion may be one reason why recreation and social activities are generally organized as nonprofit "clubs."[9] Choosing an optimal membership size is quite similar to allocating commuters to travel modes: in both cases the problem is one of restricting access to a fixed factor, which may be a golf course or a congenial social set as well as space on a highway. If it is impossible to charge higher prices to users with greater aversion to "congestion," elites may be swept out of their preserves even though they would be willing, in an all-or-nothing sense, to pay for them. Of course, the more homogeneous the elite the less the need for discriminatory pricing, and similarity of tastes is the basis for collective consumption.

---

[7] The second integral, when $\mathrm{t}_2'(L\text{-}N) > 0$, represents the value of travel time *saved* for rail commuters.

[8] The first-order condition (23) may be satisfied at $N^* < N_c$, in which case total costs have to be compared directly to find a global minimum.

[9] Avoiding income taxes and Civil Rights laws are probably more important reasons. Buchanan (1965) emphasizes decreasing average cost as the crucial factor in determining optimal club size.

APPENDIX A

Marginal Revenue with Heterogeneous Values of Time

The marginal revenue function is discontinuous at $N=\hat{N}$ unless $\hat{N}=L/2$ or $V'(\hat{N})=0$. As N approaches $\hat{N}$ from below, marginal revenue at $\hat{N}$ equals $1-\hat{N}V(\hat{N})\mathrm{t}'(\hat{N})\equiv G_1(\hat{N})$, whereas marginal revenue at $\hat{N}$ equals $1-\hat{N}V(L-\hat{N})\mathrm{t}'(\hat{N})\equiv G_2(\hat{N})$ when N approaches $\hat{N}$ from above. When $\hat{N}>L/2$, $G_2(\hat{N})<G_1(\hat{N})$. In such a case if $G_1(\hat{N})<0$, $G_2(\hat{N})$ will also be negative; if $G_1(\hat{N})>0$, $G_2(\hat{N})$ may be either positive or negative. When $\hat{N}<L/2$, on the other hand, $G_1(\hat{N})<G_2(\hat{N})$. If the first expression is positive, the second will also be positive; if $G_1(\hat{N})<0$, $G_2(\hat{N})$ may be either positive or negative.

The possibilities are enumerated below:

$\hat{N}=L/2$	Case 1: continuous marginal revenue curve
$\hat{N}>L/2$	Case 2: $G_1(\hat{N})<0$ Case 3: $G_1(\hat{N})>0$; $G_2(\hat{N})>0$ Case 4: $G_1(\hat{N})\geqq 0$; $G_2(\hat{N})<0$
$\hat{N}<L/2$	Case 5: $G_1(\hat{N})>0$ Case 6: $G_1(\hat{N})<0$; $G_2(\hat{N})<0$ Case 7: $G_1(\hat{N})<0$; $G_2(\hat{N})>0$

Since maximizing profits is equivalent to maximizing total revenue, $T(N)N$, the monopolist will set a toll at which marginal revenue is zero. This operation is straightforward if $\hat{N}=L/2$ (Case 1). Because of the discontinuity at $N=\hat{N}$, marginal revenue may be zero at two values for N when $\hat{N}\neq L/2$. This possibility is represented by Case 7, in which case total revenue at these two values for N must be compared directly. Case 4 indicates that an equilibrium at a toll of one dollar and $N=\hat{N}$ is also possible. Otherwise we will have:

$$\overline{N}V(\overline{N})\mathrm{t}'(\overline{N})-\overline{N}V'(\overline{N})[1-\mathrm{t}(\overline{N})]$$
$$=-\overline{N}T'(\overline{N})=T(\overline{N})>1$$

and

$$T'(N)N+T(N)<0$$
$$\text{for } N>\hat{N},\quad \text{Cases 2 and 6;}$$

or

$$T'(N)N+T(N)>0 \qquad \text{for } N<\hat{N}$$

and

$$\underline{N}V(L-\underline{N})\mathrm{t}'(\underline{N})$$
$$+\underline{N}V'(L-\underline{N})[1-\mathrm{t}(\underline{N})]$$
$$=T(\underline{N})<1,\quad \text{Cases 3 and 5.}$$

APPENDIX B

Example 1: Monopoly traffic flow less than Pareto optimal flow

Suppose that

$$\text{(1a)}\qquad \begin{aligned} \mathrm{t}(N)&=b, && 0<b<1,\\ & && 0\leqq N\leqq F\\ \mathrm{t}(N)&=be^{c(N-F)}, && c>0,\quad N\geqq F \end{aligned}$$

and

$$\text{(2a)}\quad V(N)=Ve^{-gN},\qquad g>0,\quad 0\leqq N\leqq L$$

By making c and g large enough we can assure that both the profit maximizing and Pareto optimal traffic flows are less than $\hat{N}=F+[1/c]\ log\ 1/b$. Denote the former by $\overline{N}$ and the latter by N^*.

N^* will exceed F if

$$\text{(3a)}\qquad \mathrm{t}'(F+0)\int_0^F V(j)dj<$$
$$1+V(F)[1-\mathrm{t}(F)]=T(F)$$

It is clear that N^* cannot be less than F because in that region $\mathrm{t}'(N)=0$. $\overline{N}$ will be less than F if

$$\text{(4a)}\qquad FV(F)\mathrm{t}'(F)-FV'(F)[1-\mathrm{t}(F)]$$
$$>1+V(F)[1-\mathrm{t}(F)]=T(F)$$

Substituting (1a) and (2a) into (3a) and (4a) we find that these inequalities are satisfied if $g=0.1$, $c=0.2$, $F=11$, $b=0.1$, and $V=150$. Therefore, for these parameter values $\overline{N}<F<N^*<\hat{N}$.

Example 2: Monopoly traffic flow greater than Pareto optimal flow

Suppose that

$$\text{(5a)}\quad \mathrm{t}(N)=b,\qquad 0<b<1,\quad N\leqq F$$
$$\mathrm{t}(N)=b+c(N-F),\qquad N\geqq F$$

$$\text{(6a)}\quad \begin{aligned} V(N) &= A - gN, \qquad N \leqq F \\ &= a, \qquad a < A, \qquad L \geqq N \geqq F \end{aligned}$$

Equations (5a) and (6a) yield the following demand curve:

$$\text{(7a)}\quad \begin{aligned} T(N) &= 1 + (A - gN)(1 - b), \; N \leqq F \\ T(N) &= 1 + a[1 - b - c(N - F)], \\ & \qquad N \geqq F \end{aligned}$$

Since $t'(N) = 0$ for $N < F$ the Pareto optimal traffic flow, N^*, cannot be less than F. According to (9) N^* will not exceed F if

$$\text{(8a)}\quad t'(F + 0) \int_0^F V(N)dN = c \int_0^F (A - gN)dN > 1 + a(1 - b)$$

Making the substitution $gF = A - a$ derived from (6a), (8a) implies that

$$\text{(9a)}\quad cF\left(\frac{A + a}{2}\right) > 1 + a(1 - b)$$

A monopolist will maximize profits by setting a toll at which marginal revenue is zero. We assume that marginal revenue is positive at $N = F - 0$, which means that

$$\text{(10a)}\quad \begin{aligned} &1 + A(1 - b) - 2g(1 - b)F = \\ &\quad 1 + A(1 - b) - 2(1 - b)(A - a) > 0 \end{aligned}$$

Marginal revenue is also assumed to be positive at $N = F + 0$, which means that

$$\text{(11a)}\quad 1 + a(1 - b) - acF > 0$$

Marginal revenue equals zero at

$$\text{(12a)}\quad \overline{N} = \frac{1 + a(1 - b + cF)}{2ac},$$

which we assume to be less than $\hat{N} = (1 + cF)/c$. Note that

$$T(\overline{N}) = \frac{1 + a(1 - b + cF)}{2a}$$

$$t(\overline{N}) = b + \frac{1 + a(1 - b - cF)}{2a}$$

The four constraints of our model are summarized below:

$$\text{(13a)}\quad F > \frac{1 + a(1 - b)}{c\left(\frac{A + a}{2}\right)} \text{ from (9a)}$$

$$\text{(14a)}\quad 1 + 2(1 - b)a > (1 - b)A \text{ from (10a)}$$

$$\text{(15a)}\quad a < \frac{1}{b + cF - 1} \text{ from (11a)}$$

$$\text{(16a)}\quad a > \frac{1}{1 + cF - b} \text{ from } \overline{N} < \hat{N}$$

Values for A, a, b, c, F which satisfy (13a)–(16a) will cause $\overline{N}$ to exceed $F = N^*$. The values $A = 1.5$, $a = 0.5$, $b = 0.1$, $c = 0.1$, and $F = 20$ satisfy these inequalities, and imply that $T(\overline{N}) = 1.225$, $t(\overline{N}) = .55$, $\overline{N} = 24.5$. Total revenue is more than \$1 higher at $\overline{N}$ than at F (\$30.13 versus \$29). On the other hand, consumer surplus plus toll revenue at F is \$38, as opposed to only \$32 at $\overline{N}$.

REFERENCES

M. Beckmann, C. B. McGuire and C. Winsten, *Studies in the Economics of Transportation,* New Haven 1956.

J. M. Buchanan, "External Diseconomies, Corrective Taxes, and Market Structure," *Amer. Econ. Rev.,* Mar. 1969, *59,* 174–77.

———, "An Economic Theory of Clubs," *Economica,* Feb. 1965, *32,* 1–14.

F. Knight, "Some Fallacies in the Interpretation of Social Cost," reprinted in G. J. Stigler and K. Boulding, eds., *Readings in Price Theory,* Vol. VI, Homewood 1952, 160–79.

A. C. Pigou, *The Economics of Welfare,* London 1918.

A. A. Walters, *The Economics of Road User Charges,* Baltimore 1969.

External Diseconomies in Competitive Supply

By CHARLES J. GOETZ AND JAMES M. BUCHANAN*

Economists have arrived at an understanding of how the competitive process eliminates the influences of irrelevant or pecuniary externalities on resource allocation.[1] They have not developed a comparably satisfactory understanding of how competition incorporates the influences of Pareto-relevant or technological externalities on allocative outcomes.[2] This paper aims to contribute toward such an understanding.

Firms in competitive industries generally impose externalities on each other through interdependencies in their input or output prices and/or through alterations of the technological production relationships. We limit discussion to externalities in supply. We shall not consider the reciprocal externalities that firms impose on each other through influences on output prices. As a second major restriction, we assume that all externalities are contained within a well-defined industry category. Finally, we limit the specific analysis to external diseconomies, although much of the analysis is fully symmetric with respect to external economies.

We argue that familiar descriptions of the nonoptimality properties of competition in the presence of technological external diseconomies are misleading in that they focus exclusively on exchange inefficiency and neglect production inefficiency. A subsidiary purpose of the paper is to suggest that only through a more detailed understanding of the way that supply externalities affect allocative outcomes can the efficacy either of organizational alternatives such as centralized industry management or of corrective devices such as taxes be considered with sufficient sophistication to yield any hope for reasonably efficient results. When the precise way in which externalities enter the competitive process is clarified, the deficiencies in the standard corrective prescriptions such as Pigovian tax schemes become evident.

I. Optimality Properties—Pecuniary Diseconomies

We shall limit analysis to a two-good, two-industry economy.[3] Consider an amended version of Adam Smith's beaver-deer model in which each of two goods is

---

* Virginia Polytechnic Institute and State University. We are indebted to several critics at various stages of this paper. We should especially mention the help provided by Leland S. Case, Milton Kafoglis, H. E. Overcast, J. A. Stockfisch, and Gordon Tullock.

[1] The famous neoclassical controversy on this point commanded the attention of such neoclassical "greats" as Alfred Marshall, A. C. Pigou, Allyn Young, D. H. Robertson, and Frank H. Knight. The central portions of the neoclassical controversy are reprinted in papers selected for inclusion in *Readings in Price Theory*, Part II. In particular, the paper by Howard Ellis and William Fellner, first published in 1943, provides an excellent summary treatment of the basic issues.

[2] Modern price theorists have recognized the complexities that remain in the analysis, and they have made progress in clarification. Specifically, note should be taken here of the contributions by E. J. Mishan, J. A. Stockfisch, and D. A. Worcester, Jr.

[3] It is not fully legitimate to generalize a two-good model into one that is applicable for the n-good economy by treating one of the two goods as a composite measure for the $n-1$ other goods. This familiar practice may be helpful in certain instances, despite the fundamental analytical error that is involved. The problem arises because of the nonuniqueness of the mix in the bundle of goods that the numeraire embodies. Once we move beyond a two-good setting, in the strictly physical sense, objective trade offs between goods in production cannot be divorced from evaluations.

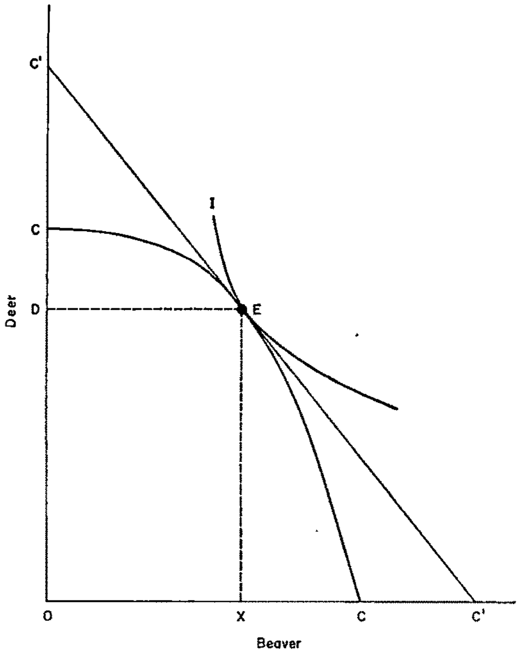


FIGURE 1

produced by a fully competitive industry. In Figure 1, the convex set of points, *CC*, is defined in the two physical dimensions, beaver being measured along the abscissa and deer along the ordinate. This curve is the locus of all possible equilibrium positions, similar to the one shown at *E*, which may be attained as demand shifts take place. Given any configuration of demand, equilibrium is reached when the market price of each good reflects both marginal cost to the producing firm and the marginal rate of substitution in consumption or use for all buyers (as reflected in the slope of the community indifference curve of Figure 1).

In the conditions depicted, each of the goods is produced under conditions of increasing supply price or average cost. At each position of full equilibrium, the familiar conditions that must be satisfied are:

$$p_i = mc_i = ac_i, \tag{1}$$

where the subscript i refers to the individual firm, and p, mc, and ac refer to output price, marginal cost, and average cost. The firm makes no profit or loss and there is no incentive for entry or exit of firms from the industry. Since the industry is merely the sum of all the firms that compose it, we also know that,

$$ac_i = ac_I \tag{2}$$

where the subscript I refers to the industry as a whole.

The construction of Figure 1 allows the function of rents in the general equilibrium context to be depicted readily. As demand patterns change, competitive adjustments have the effect of continuously converting the convex set of points traced out of *CC* into a series of linear price lines. At the equilibrium shown at *E*, the market price of beaver, in deer units, is represented by the slope of *C'C'*. If evaluated at this price, the supply of beaver, *OX*, is worth *C'D* in units of deer. This is the total industry cost or outlay that is required to produce the *OX* units of beaver. This total cost or outlay may be broken down into its two component parts, rents and nonrents. The amount *C'C* measures rents, and *CD* measures nonrents. Note that, at *E*, the average cost to the industry, and hence to the individual firm, is represented by $C'D/OX$, which is equal to market price.

To this point we have deliberately refrained from making statements about the optimality properties of the economy represented. The conditions specified in (1) and (2) are those that are necessary for a position to qualify as one of full competitive equilibrium for an industry. These may or may not be consistent with those that must be met for Pareto optimality.

When there are no technological externalities among the separate firms in the industry, the increasing supply price for the industry, depicted by the convex set of possible equilibrium points, *CC*, arises solely because of the emergence of rents.

The rents may reflect differential resource specialization, but this need not be present.[4] Rents take the form of increases in the prices of some inputs as industry output expands. Rents reflect pecuniary external diseconomies because the firm's influence on input prices is neglected. In expanding its own output, disregarding its own influence in driving up input prices, a firm imposes perfectly genuine external costs on all other firms in the industry. The error that both Marshall and Pigou made involved their interpretation of these external costs to firms, and internal to the single industry, as *net* social costs which were properly to be included in determining allocative outcomes. This error was corrected by Allyn Young, D. H. Robertson, and Frank Knight. The correction involved the demonstration that these external costs suffered by firms in the industry whose output expands are precisely offset by the receipt of income gains to the owners of inputs. Therefore, although these rents are necessary costs to the firms and to the industry taken as a whole, they are, in the terminology of modern externality theory, Pareto-irrelevant.[5] In its failure to internalize these strictly pecuniary externalities, competition guarantees overall efficiency in resource use.

These generalizations may be discussed in terms of the geometrical representation. In the absence of technological externalities, the long-run supply curve, is the *marginal social cost* curve. In Figure 1, the slope of the curve CC at any point is the marginal rate of transformation between the two goods for the society as a whole. This curve is the physically determinate *production-possibility* frontier as well as the locus of competitive equilibrium positions. Note that, at equilibrium, which under the depicted demand configuration is located at E, all of the required conditions for Pareto optimality are satisfied. The marginal rate of transformation in production equals the marginal rate of substitution in consumption or use. Market price equals marginal social cost, which is also marginal cost to the producing firms in both industries of the model. Both industries are in long-run equilibrium in that there is no incentive for resource services to be shifted. All supply decisions are made by firms, and not by industries as units.

II. Technological External Diseconomies

It is widely acknowledged that the necessary marginal conditions for Pareto optimality are not satisfied when the external effects in production are technological, that is, when the action of one firm actually modifies the physical production relationships of other firms. To qualify as a position of competitive equilibrium, however, the conditions depicted in Figure 1 must be met. As demonstrated, price must equal average as well as marginal cost for each firm in each industry. The implications of these conditions appear to have been overlooked by many writers when they introduce technological externalities. The existence of such diseconomies is commonly considered to produce a nonoptimal equilibrium position at a point such as that shown at E^* on Figure 2 with E remaining the Pareto-efficient outcome as in Figure 1.[6] If beaver are produced under conditions of external diseconomies, each firm's marginal cost is less than the actual marginal cost to society in terms of foregone real opportunities. Hence, the effects of competition would seem to be those of shifting the equilibrium away from the potential

---

[4] On this, see Joan Robinson and E. J. Mishan.

[5] This terminology is fully developed in Buchanan and Stubblebine.

[6] Perhaps the most explicit examples of this reasoning as applied within an exact analogue to our Figure 3 are found in the international trade literature, including the well-known article by Gottfried Haberler, p. 236 and p. 237n. Also see J. Bhagwati and V. K. Ramaswami, pp. 45–46.

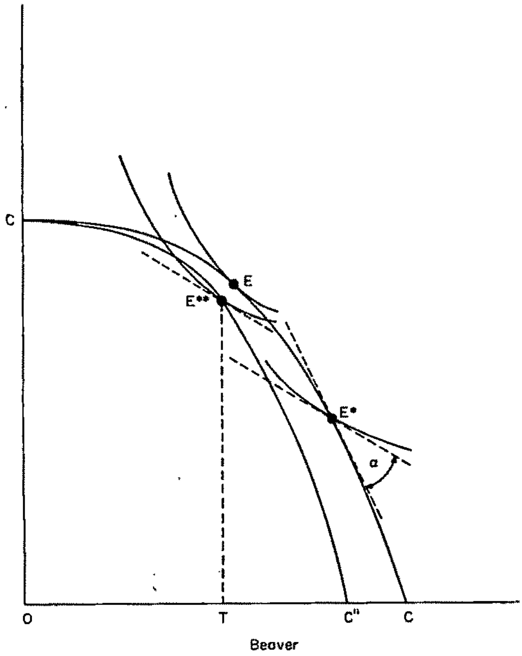


FIGURE 2

optimal point at E and along the production-possibility frontier to such a position as E^*, where there are relatively too many beaver produced. From the geometrical configuration depicted, if E^* should be the location of the inefficient equilibrium it would seem that an efficacious policy prescription might be the levy of a corrective Pigovian tax designed to eliminate the divergence, as represented by angle α, between the firm's and society's marginal rate of transformation.[7]

When the required conditions for full competitive equilibrium are incorporated into the analysis, however, the existence of such a position as E^* can be shown to be impossible. Indeed, if technological externalities exist, competitive equilibrium must be located on some *interior* locus, such as CC''. Production inefficiencies in addition to the exchange inefficiencies suggested in the more orthodox treatments must characterize competitive equilibrium under these conditions. The next step is that of indicating precisely how these production inefficiencies emerge.

Assume that a full competitive equilibrium is established at a beaver output of T units in Figure 2 and that the conditions for equilibirum are fully satisfied. Our problem is one of demonstrating that, with technological externalities, T is produced at E^{**} inside rather than on the *efficient* production-possibility frontier, CC. Since their existence generates somewhat different results, we shall distinguish two types of external effects on production: output-generated and input-generated.[8] Output-generated externalities occur when a firm's technological production function contains the output of other producers as one of its arguments. The original neoclassical discussion seems to be in these terms and this case involves a strict interpretation of the phrase "external economies of industry [output] scale." Input-generated externalities occur when the externality argument in a firm's production function is based on other firm's use of one or more particular inputs. We will deal with these two cases successively.

Output-Generated Externalities

Our demonstration involves examining the behavior of an individual firm in the affected industry. In order to emphasize the physical input-output relationship, cost is measured in units of a single homogeneous input or an input bundle whose use is to be minimized.[9] For simplicity, all firms are regarded as identical. Assume, then, that the cost function of any firm has the following specific functional form:

---

[7] Bhagwati and Ramaswami prescribe such a corrective tax for the analogous case of an external economy.

[8] Murray Kemp provides a separate analysis of these two classes of externalities. However, as indicated in the text below, our own analysis generates different conclusions.

[9] Conversion of cost into value terms requires only that we multiply both sides of (3) by input price. We neglect this complication since cost minimization for the firm implies an identical solution in either case. For the *industry*, cost minimization in value terms is clearly inappropriate since this would involve internalization of purely pecuniary, nonrelevant externalities.

$$(3)\quad c_i = k + q_i^a + Q_i^b q_i \qquad a > 1,\ b > 0$$

where q_i is the firm's own output, where Q_i is the aggregate output of all *other* firms in the industry, and where $b>0$ indicates the presence of the technological external diseconomy. As noted in (1) above, full competitive equilibrium requires that each firm operate where $mc_i = ac_i$, an output that corresponds to the minimum point on the firm's average cost curve. This equilibrium is depicted by point e_i in Figure 3. Average and marginal cost are derived straightforwardly from (3) as,

$$(4)\qquad ac_i = kq_i^{-1} + q_i^{a-1} + Q_i^b,$$

and

$$(5)\qquad mc_i = aq_i^{a-1} + Q_i^b$$

with the externality variable, Q_i, regarded as a parameter by the firm. With all firms operating at a scale q_i, which equates (4) and (5), the industry produces the output T with n firms[10] such that

$$(6)\qquad n = T/q_i$$

The industry can, however, be reorganized with fewer firms each of which produces a larger output. Such a reorganization will lower average cost for each firm in the industry, and, hence, a smaller quantity of industry inputs will be required to produce the same output, T.

In order to demonstrate this production-inefficiency of the competitive outcome, it is necessary to replace the *ceteris paribus* cost curves faced by the individualistic firms with *mutatis mutandis* cost curves which reflect Q_i as a variable rather than as a parameter.[11] The "other firms' output" externality variable is,

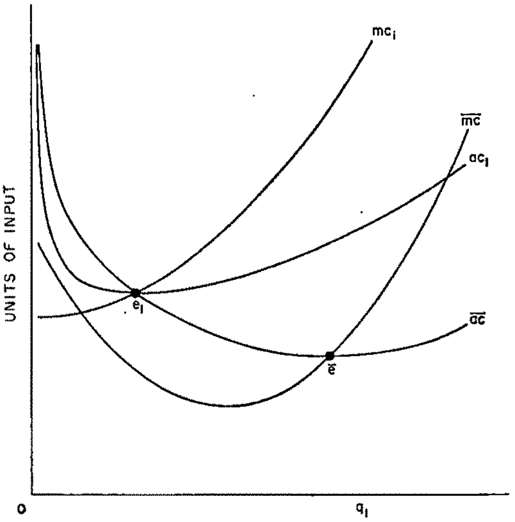


FIGURE 3

$$(7)\qquad Q_i = T - q_i,$$

for any constant industry output, T. Substituting (7) into (3) for Q_i, we get a *mutatis mutandis* total cost function,

$$(3b)\qquad \bar{c} = k + q_i^a + (T - q_i)^b q_i,$$

where industry output is held constant at T by varying n inversely with q_i, according to the relationship implied in equation (6). The firm's *mutatis mutandis* average and marginal cost functions are:

$$(8)\quad \overline{ac} = kq_i^{-1} + q_i^{a-1} + (T - q_i)^b,$$

$$(9)\quad \overline{mc} = aq_i^{a-1} + (T - q_i)^b - q_i b(T - q_i)^{b-1}$$

The average costs in (4) and (8) are related as shown in Figure 3, with $ac_i > \overline{ac}$ for expansions and $ac_i < \overline{ac}$ for contractions from the firm's competitive equilibrium where $ac_i = mc_i$. Significantly, $\overline{ac}$ is not minimized where $ac_i = mc_i$ because, as (5) and (9) show, mc_i is always greater than

---

[10] We adopt here the convenient expositional fiction that the number of firms is a continuous variable. In actuality, of course, the solution in the text must be modified to allow for n being always an integer. If n is large, this adjustment becomes negligible.

[11] Failure to recognize this sort of distinction seems to have led Kemp to reach conclusions quite different from those we reach in this section.

$\overline{mc}$. The total resource cost of producing T units is minimized when firms produce at $\bar{e}$, the low point of $\overline{ac}$ in Figure 3, that scale of output which equates (8) and (9). Since (4) and (5) must be equalized for competitive equilibrium and since (8) and (9) must be equalized for Pareto optimality, it is evident that competition is production-inefficient in that firm sizes are suboptimal and that there are too many firms in the industry.[12] Competitive equilibrium necessarily lies on an interior locus such as that shown by CC'' in Figure 2.

Our example may now be framed in completely general terms. Assume that the firm's total cost c_i is a function of its own output and external output

$$c_i = f(q_i, Q_i) \tag{10}$$

Then the $ac = mc$ condition of competitive equilibrium requires

$$\frac{c_i}{q_i} = \frac{\partial f}{\partial q_i} \tag{11}$$

But if industry output T has been constrained to be constant by equation (8) above, efficient firm operation demands that

$$\frac{c_i}{q_i} = \frac{\partial f}{\partial q_i} + \frac{\partial f}{\partial Q}\frac{\partial Q_i}{\partial q_i} = \frac{\partial f}{\partial q_i} - \frac{\partial f}{\partial Q} \tag{12}$$

for each firm.

An extremely important implication of this analysis is that any attempt to impose a per unit corrective tax on the output of firms in the industry will fail. Such a tax will serve only to move the equilibrium to a different point on the inefficient interior locus. A per unit tax would simply shift the curves in Figure 3 upward in a parallel fashion. Or, in terms of the equations (11) and (12), adding a corrective tax t to each cost concept does nothing to modify the basic inconsistency of their solutions. Any fully corrective tax must be more complex than has been realized, since not only marginal cost but also the firm's perception of the *shape* of its average cost curve must be modified.[13]

Input-Generated External Diseconomies

The analysis above has been confined to "output-generated" external diseconomies. Similar conclusions are applicable for "input-generated" external diseconomies, those which are directly related to the use of one or more industry inputs and through these indirectly to the actual scale of industry output. In this case, the elements of our analysis are based upon a familiar adaptation of the Edgeworth-Bowley box construction to derive an economy's efficient transformation locus. To our knowledge, however, the implications have not been explicitly applied to externalities.

Wolfgang Stolper and Paul Samuelson used the box construction to demonstrate that the ratios of the marginal products of any two factors must be identical in each industry if the economy is to be production-efficient. Competitive equilibrium implies the equalization of these ratios of marginal products to the *firm*. But the Stolper-Samuelson condition must be interpreted in terms of social marginal product, that is, *net* marginal product to the industry. Since these two concepts of marginal product diverge when technological externalities exist, it seems clear that, in the standard case where factor prices are identical to all industries, input-generated external diseconomies lead to interior production-inefficient competitive equilibria. Specifically, the social marginal physical product of a diseconomy-producing input will be excessively low relative to the marginal physical products of other inputs. Consequently, such an input is over-used in the industry where it produces the diseconomies. A particularly clear illustration of this is provided when there exists an input whose marginal product to the firm is positive but whose externality-producing impact on the production of other firms causes its net (industry) marginal product actually to become negative.

---

[12] As the geometrical construction suggests, the nonoptimality in firm size reflects the presence of an incentive for industry-wide internalization of decisions that is absent in the orthodox competitive equilibrium. Such an incentive always exists with respect to potential exploitation of monopoly-monopsony gains, but in the external diseconomy model with firm-size inefficiency there is some incentive over and beyond this. However, to the extent that individual firms behave parametrically, there will be no more tendency to depart from the equilibrium established in the external diseconomy case than there is in the more familiar model. In the absence of transactions costs, the inefficiency reflected by nonoptimal firm sizes could not, of course, remain.

[13] We hope to explore some of the complex issues involved in the determination of corrective taxes in a separate paper in process. We should note that our implied criticism of the simple corrective tax is different from, although related to, that of Davis and Whinston. Their analysis was developed exclusively in a small-number setting, and they did not consider the effects of the constraints imposed by the overall conditions for competitive equilibrium.

The input-generated externalities are incorporated in inefficient input combinations for the firm regardless of the outputs chosen. It is not possible to draw conclusions about firm size similar to those previously stated with respect to output-generated externalities. With input-generated external diseconomies, firms may be either too small, the right size, or too large. The determination of efficiency-producing corrective taxes remains, however, a more complicated problem than has been generally acknowledged.[14]

IV. Conclusions

Economies of space have prevented us from relating our analysis more directly to the neoclassical discussion and also from following up several of the implications in detail. Our objective here has been limited to presenting the analytical foundations for the theory of supply in environments that embody actual or potential external diseconomies. We have made no effort to provide a full taxonomy even of this subset of externalities. Specifically, we have restricted consideration to external diseconomies which are contained within a well-defined industry.

In this context, our results indicate that the standard description of misallocation in the presence of external production diseconomies is misleading.[15] The impact of production externalities has heretofore been conceived of exclusively in terms of a suboptimal exchange equilibrium on the production possibilities frontier. We have shown, however, that this equilibrium must take place on a pseudo-frontier which lies interior to the efficient production possibilities locus. The resulting combination of exchange-inefficiency with production-inefficiency renders the construction of corrective devices much more difficult.

It is also possible that the output of a good produced under external diseconomies will actually be less than the Pareto-optimal quantity, in contrast to the standard expectation that such a good is always over-produced. It should be apparent from the construction of Figure 3 that utility functions embodying the standard properties could produce an externality-distorted competitive equilibrium that falls either to the right or to the left of the optimal quantity shown at *E*. The directional relationship between the competitive industry output and the socially optimal output depends strictly on the particular characteristics of utility functions as well as the strength of the production externality component.

---

[14] One aspect of the difficulties involved in the correction of input-generated externalities has been described by Charles Plott.

[15] The paper by John S. Chipman appeared only after the initially submitted version of this paper was provisionally accepted for publication. Accordingly, we have not modified the text to cover points raised by Chipman. His paper seems to be a sophisticated and modern presentation of the orthodox analysis. Our results differ from those reached by Chipman primarily because Chipman treats each firm's production function as a parametric approximation to the industry production. This simplification has unintended results. For instance, it does not generate U-shaped firm cost curves, nor does it allow the industry's resource cost to be affected by the distribution of output among firms.

In summary, our analysis suggests that no real paradigm has existed in this area of the theory of competitive supply. To our knowledge, no one has explicitly traced the impact of externalities through their influence on the individual competitive firm's necessary conditions for competitive equilibirum. This approach has enabled us to describe more adequately the precise nature of a nonoptimal equilibrium in the presence of one form of technological externality.

REFERENCES

K. J. Arrow and T. Scitovsky, eds., *Readings in Welfare Economics*, Homewood 1969.

J. Bhagwati and V. R. Ramaswami, "Domestic Distortions, Tariffs, and the Theory of Optimum Subsidy," *J. Polit. Econ.*, Feb. 1963, *71*, 44–50.

J. M. Buchanan and W. C. Stubblebine, "Externality," *Economica*, Nov. 1962, *29*, 371–84.

J. S. Chipman, "External Economies of Scale and Competitive Equilibrium," *Quart. J. Econ.*, Aug. 1970, *84*, 347–85.

O. A. Davis and A. Whinston, "Externalities Welfare and the Theory of Games," *J. Polit Econ.*, June 1962, *70*, 241–62.

H. S. Ellis and W. Fellner, "External Economies and Diseconomies," *Amer. Econ. Rev.* Sept. 1943, *33*, 493–511; reprinted in G. J Stigler and K. E. Boulding, eds., *Readings i Price Theory*, 242–63.

G. Haberler, "Some Problems in the Pur Theory of International Trade," *Econ. J.* June 1950, *60*, 233–40.

M. Kemp, "The Efficiency of Competition a an Allocation of Resources; Part I," *Can. J Econ.*, Feb. 1955, *21*, 30–42.

E. J. Mishan, "What is Producers' Surplus?,' *Amer. Econ. Rev.*, Dec. 1968, *58*, 1269–82

C. Plott, "Externalities and Corrective Taxes,' *Economica*, Feb. 1966, *33*, 84–87.

J. Robinson, "Rising Supply Price," *Economica*, 1941, *8*, 1–8; reprinted in *Readings i Price Theory*, pp. 233–41.

G. J. Stigler and K. E. Boulding, eds., *Reading in Price Theory*, Homewood 1952.

J. A. Stockfisch, "General Equilibrium versu Partial Equilibrium Supply Curves," unpublished mimeo. paper, 1962.

———, "A Study in the Theory of Cost an Supply," mimeo paper, Institute for Defens Analyses, 1970.

W. F. Stolper and P. A. Samuelson, "Protection and Real Wages," *Rev. Econ. Stud.* Nov. 1941, *9*, 58–73.

D. A. Worcester, Jr., "Pecuniary and Technological Externality, Factor Rents and Social Costs," *Amer. Econ. Rev.*, Dec. 1969 *59*, 873–85.

Factor Substitution and Effective Protection Reconsidered

By MORDECHAI E. KREININ, JAMES B. RAMSEY, AND JAN KMENTA*

Numerous articles have appeared in recent years concerning the theory and estimation of the effective protective rate (see, for example, W. M. Corden, Bela Balassa, Giorgio Basevi, J. Clark Leith (1971)), defined as the amount of protection accorded by the tariff structure to the domestically produced value added. This theory is then utilized, among other things, to draw conclusions concerning the reactions of the economy to tariff changes and their welfare implications (see Harry Johnson, Leith and G. L. Reuber, and Balassa, Stephen Guisinger, and Daniel Schydlowski). Yet, in the received body of economic theory such inferences are normally drawn from the profit maximizing behavior of the firm and the utility maximizing behavior of the consumer. While the consumer reacts to price changes on the final product (and is therefore affected by nominal and not by effective protection), the producer's action is determined by the prices of both the final product and the inputs. This, however, is not equivalent to saying that the producer reacts to changes in value added. Rather, economic theory suggests that he would react to changes in profit. If the producer is assumed to combine local labor, l (the primary input), with an imported input, x, to produce his output, q, then domestic value added is comprised of profit plus the cost of labor input. Consequently, predictions as to the changes in resource allocation based on changes in profit and in value added will lead to the same conclusions only in special circumstances; most notably in the case of fixed-coefficient homogeneous first degree production functions where output is determined by resource input constraints.

The concept of effective protection deals with the protection accorded by the tariff structure to wages and profits combined. Alternatively, it has on occasion been defined as the protection accorded to domestic labor. Although this is a legitimate line of inquiry on its own merits, it is not that rate of protection which determines the expansion of domestic production and the consequent curtailment of imports and misallocation of resources attendant upon a rise in tariffs. Rather, it is the changes in the rate of profit that govern adjustments in domestic output. When factor substitution is permitted, the two approaches may not yield equivalent results.

This paper departs from earlier work in the field in that it examines the issues raised by effective protection in the light of neoclassical economic theory and is, therefore, built explicitly around the principle of profit maximization. In particular, we determine the short-run change in domestic output, value added, and profit, employing the *CES* production function. The changes in output and imports under the tariff structure emerge from the model, as do the changes in protection to domestic factors. One important result is that even in partial equilibrium analysis the effective protection concept as presently defined is limited in its usefulness and in some situations leads to incorrect conclusions.

---

* Professor, associate professor, and professor of economics at Michigan State University, respectively. We are grateful to P. J. Lloyd of the Australian National University for helpful comments on an earlier draft of this paper.

In the next section we analyze the effects of changes in the tariff structure on domestic output, labor input, and imports under profit maximizing behavior, with reference to the *CES* production function. The analysis is of a partial equilibrium short-run variety where the stock of capital is assumed fixed. Section II develops and compares four alternative definitions of "effective rates of protection" in both continuous and discrete forms. In particular, we examine the usefulness of the various definitions in shedding light on the change in the allocation of resources in response to a change in the tariff structure.

I. Analysis of the Profit Maximizing Model

Consider a small country importing input X and final good Q which is also produced domestically, and assume that the domestic production of Q utilizes only two inputs, the imported input (X) and local labor (L). The country is assumed to be small enough so as not to affect the terms at which it trades; i.e., it is a price taker on the international market. Also, domestic prices of all products and factors are determined competitively. In particular this assumption applies to the wage rate, w, and to the prices of goods Q and X, denoted p and r, respectively. In addition, we assume that, in the short run, the number of firms is fixed and is given as n. The last assumption enables us to determine the changes in output of the industry caused by changes in the tariff structure without getting involved in a discussion of demand in general equilibrium terms.

Finally, we assume that the marginal cost curves of the n firms are identical and that the production functions have decreasing returns to scale in the variable inputs so that the optimal size of firm as well as the number of firms is determined.[1]

The *CES* production function is given as

$$q = (\delta_1 x^{-\rho} + \delta_2 l^{-\rho})^{-\nu/\rho}, \qquad \nu < 1,$$

where, for each firm, q is output, and x and l are the quantities of the imported and labor inputs, respectively. The levels of industry output and inputs are $Q=nq$, $X=nx$, $L=nl$, where n is the number of firms. The short-run profit maximizing solutions for the equilibrium quantities are

$$(1) \qquad Q_s = \mu_1 p^{\lambda} R^{-\lambda/(1-\sigma)},$$

$$(2) \qquad X = \mu_2 p^{1+\lambda} r^{-\sigma} R^{-(1+\lambda-\sigma)/(1-\sigma)},$$

$$(3) \qquad L = \mu_3 p^{1+\lambda} w^{-\sigma} R^{-(1+\lambda-\sigma)/(1-\sigma)},$$

where $\mu_1 = n\nu^{\nu/(1-\nu)}$, $\mu_2 = n\nu^{1/(1-\nu)}\delta_1^{\sigma}$, $\mu_3 = n\nu^{1/(1-\nu)}\delta_2^{\sigma}$, $\lambda = \nu/(1-\nu)$, $R = (\delta_1^{\sigma} r^{1-\sigma} + \delta_2^{\sigma} w^{1-\sigma})$ and $\sigma = 1/(1+\rho)$.

Further, the post-tariff prices of the output and of the imported input are $p = p_o(1+t)$ and $r = r_o(1+m)$, respectively, where t and m are the tariff rates and p_o, r_o are the pre-tariff (free trade) prices.

1. *Changes in the Tariff Structures and Domestic Production*

In order to examine the effect on domestic production of simultaneous changes in the tariff on the output and on the input—t and m, respectively—we take the total differential of the output equation (1) with respect to t and m to obtain

$$(4) \quad dQ = \lambda Q_s \left[\frac{dt}{(1+t)} - g \frac{dm}{(1+m)} \right],$$

where $g = (r^{1-\sigma}\delta_1^{\sigma})/R$ and where by definition $0 < g < 1$. It can be shown by substitution of the equilibrium values of the variables that $g=rX/(rX+wL)$, i.e., g is the ratio of import cost to total cost.

---

[1] The complete construction of the model and the explicit derivation of the supply equations from the specification of the production functions are not shown here in the interest of conserving space. They are available from the authors upon request.

The change in t which exactly offsets a given change in m leaving domestic output unchanged is obtained by setting $dQ=0$. In the Cobb-Douglas case, i.e., when $\sigma=1$, this leads to

$$dt = \frac{(1+t)}{(1+m)} \frac{\alpha_1}{\alpha_1+\alpha_2} dm, \tag{5}$$

where α_1 is the exponent of x and α_2 is the exponent of l in the Cobb-Douglas production function. Thus, if initially $t=m$ (and as a special case $t=m=0$), then a given change in m must be accompanied by a fraction $\alpha_1/(\alpha_1+\alpha_2)<1$ of that change in t in order to leave domestic output unaffected. To achieve a reduction in domestic output under the same conditions, one requires $dt<\alpha_1/(\alpha_1+\alpha_2)dm$. Alternatively, output will decrease with an equal change in t and m if $(1+t)/(1+m) > (\alpha_1+\alpha_2)/\alpha_1$. Also, the closer firms are to constant returns to scale, i.e., the closer $(\alpha_1+\alpha_2)$ is to one, the greater the change in Q in response to any equal change in tariff rates. Finally, the greater the share of the imported input in total cost, the smaller the increase in Q resulting from an equal change in both tariffs.

The results for $\sigma\neq 1$ are similar. The major difference is that in the latter case the critical values for dt and dm which yield decreases in equilibrium output depend upon the domestic input prices, whereas in the Cobb-Douglas case the solution is invariant with respect to input prices. This difference is a result of the nonunitary elasticity of substitution between the input factors in the *CES* case.

2. *Changes in Tariff Structure and the Equilibrium Level of Labor Input*

The effect of changes in the tariff structure on the equilibrium level of employment (L) can be determined by obtaining the total differential of the labor equation (3):

$$dL = \frac{L}{1-\nu}\left[\frac{dt}{(1+t)} - (1-\sigma+\sigma\nu)g\frac{dm}{(1+m)}\right] \tag{6}$$

where $g=(r^{1-\sigma}\delta_1^{\sigma})/R$ as in (4). From equation (6) it is clear that $\partial L/\partial t>0$. When $\sigma=1$, $\partial L/\partial m<0$; that is, the substitution effect of an increase in the domestic price of the imported input is outweighted by the scale effect, so that the overall effect of the increase in m is to reduce the equilibrium level of labor input. In general, the sign of $\partial L/\partial m$ depends, as one would expect, on the elasticity of substitution and on the degree of returns to scale. If $\sigma<1/(1-\nu)$, then $\partial L/\partial m<0$, and if $\sigma>1/(1-\nu)$, $\partial L/\partial m>0$. In the special case where $\sigma=1/(1-\nu)$, the equilibrium quantity of labor is invariant to changes in the tariff on the imported input as the substitution effect of the price change is negated by the scale effect.

3. *Changes in the Tariff Structure and Import Levels*

The effect of a change in the tariff structure on the quantity of final product that is imported is obtained by taking the total differential of $Q_i=(Q_d-Q_s)$ with respect to t and m, where Q_i, Q_d, Q_s are the equilibirum quantities imported, demanded, and supplied. As noted earlier $dQ_d/dt<0$, but dQ_s may be positive or negative. Thus, whether imports increase or decrease depends unambiguously on the sign of dQ_s. In order to have an increase in the quantity of Q imported, one must have $dQ_s<dQ_d$. However, given the preceding analysis, such a situation can occur for given dt only if dm is sufficiently large. We have already indicated that in the special case of the Cobb-Douglas production function with equal initial

tariff rates, dm must be greater than $[(\alpha_1+\alpha_2)/\alpha_1]dt$ in order to have $dQ_s<0$, where $(\alpha_1+\alpha_2)/\alpha_1>1$. Thus, we require an even larger change in m relative to that in t in order to achieve an increase in imports. If a reasonable value for the ratio $(\alpha_1+\alpha_2)/\alpha_1$ is 2.0, then for the size of changes in tariffs normally encountered in actual situations one would not expect to find an overall increase in tariff rates leading to an increase in imports. However, if at the time of the change in the tariff structure m is very small and t is very large, then only a modest difference in the size of dm and dt would suffice to lead to an increase in imports.

II. The Profit Maximizing Model and Concepts of Effective Protection

The rationale underlying the idea of effective protection is that general tariff changes shift not only the output price, but also the input prices. Consequently, the reallocation of resources resulting from a given change in tariffs depends on the simultaneous shift in the industry's product supply and input demand curves. The approach discussed in the previous sections is one way of accomplishing this; defining effective tariff rates as the protection accorded to value added is another. However, effective rates of protection were originally defined for situations considerably different from that envisioned in our analysis. Thus, the "effective tariff world" might be viewed as one in which the aggregate (industry) production function is homogeneous of the first degree, no input substitution is possible (at least in the short run), and the output level is determined by short-run supply constraints on the inputs. To see more clearly the contrast between the situation examined in this paper and that traditionally studied in the effective tariff literature, observe Figure 1. The supply curve is drawn perfectly horizontal with a maximum possible output level of Q_1, which is determined by input resource constraints. S is the supply curve before the tariff is imposed on the imported input and S^* is the supply curve after the tariff, while p is the domestic (tariff ridden) price and p_o is the free trade price. The quantities demanded domestically before and after the tariff are Q_2 and Q_4, respectively. Domestic quantity Q_1 will be supplied after the tariff is imposed provided that $p\geq S^*$. Note that $(p-S^*)$ is the (maximized) profit per unit which—under the assumption of fixed domestic wage rates—is equal to the change in "value added" per unit. Under these conditions the relative change in value added is, indeed, a reasonable measure of the short-run cost of protecting the industry and can indicate the potential long-run misallocation of resources. However, the traditional definition of effective tariff rate loses much of its appeal when one moves from the world described in Figure 1 to that described in Figure 2. Here D and S are the pre-tariff domestic demand and supply curves of the good Q, with the international price being p_o. The foreign supply curve is assumed to be infinitely elastic at the price p_o, so that imports are (Q_2-Q_1) and domestic production is Q_1. When a tariff rate t is im-

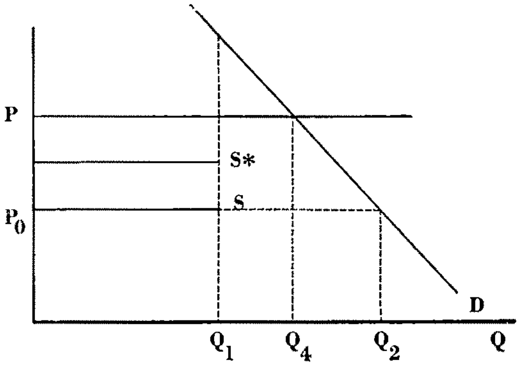


Figure 1. Domestic Demand for and Supply of Good Q in the Fixed Coefficient Homogeneous First Degree Production Function Case with Supply Constraints on Inputs

posed on Q, the domestic price of imports rises to p. A tariff rate m levied on the imported input, x, would shift upward the marginal cost of producing Q, so that the new domestic supply curve is S^*. Under the tariff structure imports become $(Q_4 - Q_3)$ and domestic producion is Q_3.

Further, it will become clear presently that attempts to "adapt" the "standard" definition of effective tariff to other types of production functions can lead to incorrect conclusions about the effects of tariff changes on resource allocation.

1. *Alternative Definitions of Effective Tariff Rate*

The conventional definition of the effective rate of protection is the proportional change in value added for a given good Q; i.e., $\tau_v^\Delta = (V' - V)/V$ where V and V' are, respectively, value added before and after the imposition of tariffs. To carry out the analysis in a manner consistent with the continuous formulation of the model used so far, let us define a continuous form of the effective rate of protection by $\tau_v = dV/V$, where dV is the total differential of value added with respect to the constituent nominal tariff rates. Thus, τ_v is the relative rate of change in value added. In the analyses below, we shall use both τ and τ^Δ.

In addition to the distinction between the continuous and discrete forms of effective tariff rates, one might consider the alternative of defining effective tariff rates with respect to variables other than value added; in particular, the variables Q_s, L, and $\Pi = n\pi$, where Π is total profit in the industry. The results can then be compared to those obtained for domestic value added. Accordingly, let us define $\tau_q = dQ/Q$, $\tau_l = dL/L$, $\tau_\pi = d\Pi/\Pi$; with the corresponding discrete effective tariff rates $\tau_q^\Delta = (Q' - Q)/Q$, $\tau_l^\Delta = (L' - L/L$, $\tau_\pi^\Delta = (\Pi' - \Pi)/\Pi$, where the primed symbols refer to the post-tariff, and the unprimed symbols to the pre-tariff situation. Under the stated conditions of the model, $\Pi = n\pi$, $V = nv = n(\pi + wl)$, $L = nl$, and $Q = nq$, where π, v, l, and q represent, respectively, profit, value added, labor input and output of the individual firm.

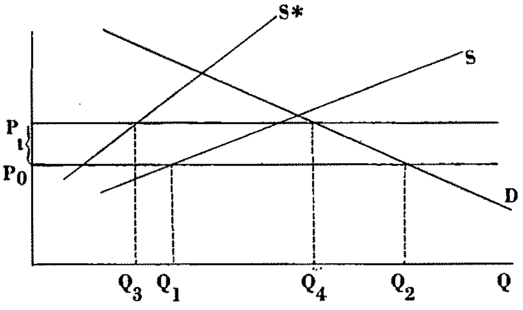


FIGURE 2. DOMESTIC DEMAND FOR AND SUPPLY OF THE GOOD Q BY THE IMPORTING COUNTRY

2. *Properties of Alternative Effective Tariff Rates*

In this subsection we consider the relationship between the various definitions and compare the results obtained for the Cobb-Douglas function with those for the general *CES* case.

Cobb-Douglas Production Function

In the Cobb-Douglas case the equilibrium conditions yield the following relations:

$$(7)\qquad rX = \alpha_1 pQ_s, \qquad wL = \alpha_2 pQ_s, \qquad \Pi = (1 - \alpha_1 - \alpha_2)pQ_s,$$

where α_1 is the imported input's share of the product, α_2 is labor's share, and $(1 - \alpha_1 - \alpha_2)$ is the share going to profit. In this situation, value added is $V = \Pi + wL = (1 - \alpha_1)pQ_s$. From equation (4) and setting $\sigma = 1$, we obtain:

$$(8)\quad \tau_q = \frac{dQ}{Q} = \left(\frac{\alpha_1 + \alpha_2}{1 - \alpha_1 - \alpha_2}\right)\left[\frac{dt}{(1+t)} - \frac{\alpha_1}{\alpha_1 + \alpha_2}\,\frac{dm}{(1+m)}\right]$$

Since τ_q and dQ are proportional, the earlier analysis of dQ (see equation (4)) holds for τ_q. Further, since Π and V are proportional to wL, we must have $\tau_\pi = \tau_v = \tau_l$. Using the above equations we find that the common value of τ_π, τ_l, and τ_v is:

$$(9) \quad \tau_v = \frac{dV}{V} = (1 - \alpha_1 - \alpha_2)^{-1} \cdot \left[\frac{dt}{(1+t)} - \alpha_1 \frac{dm}{(1+m)}\right]$$

To compare τ_q and τ_v, we rewrite τ_v as $\tau_v = \tau_q + dt/(1+t)$. Clearly, $\tau_q < \tau_v$ for all positive changes in the tariff rate. More importantly, a negative τ_v is a sufficient, but not a necessary, condition for τ_q to be negative.

The discrete formulations of the effective tariffs can be derived from the basic definitions and the use of the profit maximizing conditions:

$$(10) \quad \tau_q^{\Delta} = \left[(1+t)^{(\alpha_1+\alpha_2)}(1+m)^{-\alpha_1}\right]^{1/(1-\alpha_1-\alpha_2)} - 1$$

The necessary and sufficient condition for $\tau_q^{\Delta} \geq 0$ is

$$(1 + t) \geq (1 + m)^{\alpha_1/(\alpha_1+\alpha_2)}$$

where $\alpha_1/(\alpha_1+\alpha_2) = rX/(rX+wL) = r_oX_o/(r_oX_o+wL_o)$, $r = r_o(1+m)$; X and L are the post-tariff equilibrium input levels, and X_o and L_o are the pre-tariff equilibrium input levels. Since in the Cobb-Douglas case the ratio of the equilibrium value of imported input to total cost is a constant, equality of t and m leads to a strict inequality, so that $\tau_q^{\Delta} > 0$. To clarify the implications of this result, consider Table 1 which shows for various values of $\alpha_1/(\alpha_1+\alpha_2)$ the increases in m required to offset increases in t in order to keep Q constant.

The discrete form of the effective tariff on the labor input is obtained in a manner similar to τ_q^{Δ}:

$$(11) \quad \tau_l^{\Delta} = \left[(1 + t)(1 + m)^{-\alpha_1}\right]^{1/(1-\alpha_1-\alpha_2)} - 1$$

Clearly, $\tau_l^{\Delta} = \tau_\pi^{\Delta} = \tau_v^{\Delta}$ for the same reasons given for the continuous case. With respect to the relationship between τ_q^{Δ} and τ_v^{Δ} we see by reorganizing the expressions for τ_q^{Δ} and τ_v^{Δ} that $\tau_v^{\Delta} = (1+t)(1+\tau_q^{\Delta}) - 1$, from which we conclude that $\tau_q^{\Delta} < \tau_v^{\Delta}$ for all $t \geq 0$.

The effect of a change in the tariff structure on the quantity of the good supplied by domestic producers can be measured by τ_q or τ_q^{Δ}. The use of τ_v (or τ_v^{Δ}) has distinct shortcomings. In particular, a positive value of τ_v does not necessarily mean that the domestic quantity supplied has increased. In addition, τ_v provides no direct measure of the degree of change in Q_s as does τ_q.

The effect of a change in the tariff structure on the equilibrium quantity of the labor input can be measured by τ_l (or τ_l^{Δ}). Clearly a positive or negative value for τ_l indicates an increase or a decrease in labor input and the magnitude of τ_l is a measure of the proportional change in the labor input. However, in the Cobb-Douglas case $\tau_l \equiv \tau_v$, so that the usual definition of the effective tariff rate also measures the effect of tariff changes on the equilibirum labor input.

The effects of a change in the tariff structure on imports depend on the relative effects of the tariff change on the quantities demanded and supplied. Effective tariff rates provide at best an indirect indication of the direction of change but

Table 1—Level of Tariff on the Final Product that is Necessary to Offset the Effect on Output of a Given Tariff on Input

Tariff on Input	Import cost as a proportion of total cost		
	0.25	0.50	0.75
10%	2.5%	4.9%	7.4%
20%	4.7%	9.5%	14.5%
30%	6.8%	14.0%	21.7%
40%	8.8%	18.3%	28.7%
50%	10.7%	22.5%	35.5%

do not measure the magnitude of that change. The "natural" measure of the effect of tariffs on imports is τ_q. However, the sign of the change in the equilibrium level of imports can be unambiguously determined by the sign of τ_q only if the change in quantities demanded and supplied are in the opposite direction. Thus, if an increase in t leads to a decrease in quantity demanded and if the net effect of the increases in t and m is to increase quantity supplied ($\tau_q > 0$), then one can conclude that the level of imports will decline.

CES Production Function

The specific formulation of τ_q in the *CES* case is obtainable from equation (4), which gives the total differential of Q:

$$\tau_q = \lambda \left[\frac{dt}{(1+t)} - g \frac{dm}{(1+m)} \right], \tag{12}$$

where $\lambda = \nu/(1-\nu)$, and $g = (r^{1-\sigma}\delta_1^\sigma)/(\delta_1^\sigma r^{1-\sigma} + \delta_2^\sigma w^{1-\sigma})$ is the equilibrium ratio of the cost of import to total cost. Unlike the Cobb-Douglas case, τ_q in the *CES* case depends upon the prices of the inputs.

With the *CES* production function the equilibrium profit level is given by $\Pi = (1-\nu)pQ_s$, where Q_s is the equilibrium output of the domestic industry, and τ_π is obtained from this definition and equation (4) as

$$\tau_\pi = (1-\nu)^{-1} \left[\frac{dt}{(1+t)} - \nu g \frac{dm}{(1+m)} \right] \tag{13}$$

Comparing τ_π and τ_q, we observe that τ_q is strictly less than τ_π.

From equation (3) we derive the rate of effective protection on labor as

$$\tau_l = (1-\nu)^{-1} \cdot \left[\frac{dt}{(1+t)} - \sigma(\nu+\rho) g \frac{dm}{(1+m)} \right] \tag{14}$$

In the Cobb-Douglas case, it was observed that τ_π and τ_l were identical. The shift to a nonunitary elasticity production function removes this identity.

The expression $\sigma(\nu+\rho)$ can be rewritten as $1-\sigma(1-\nu)$. Thus, if $0 < \sigma < 1/(1-\nu)$ then the term $\sigma(\nu+\rho)gdm/(1+m)$ has the same sign as the corresponding terms for dm for τ_q and τ_π. In short, with $dt = 0$, an increase in m decreases labor input. However, if $\sigma > 1/(1-\nu)$ the term changes sign and the labor input will *increase* with an increase in m. That is, when $\sigma > 1/(1-\nu)$ the substitution effect of an increase in m outweighs the scale effect, so that the equilibirum input increases. However, labor input could still increase even if σ is inelastic, provided $dt > [(1+t)/(1+m)] g\sigma(\nu+\rho)dm$.

In conclusion, we deduce from these results that labor input decreases with an increase in the tariff on the imported input only if $\sigma < 1/(1-\nu)$ and the inequality for dt in terms of dm does not hold.

The last of the alternative definitions of effective tariff is τ_v, the explicit formula for which can be derived from equation (4) and from the identity $V = pQ_s - rX$. After some algebraic manipulation and noting that $rX/pQ_s = \nu g$, we obtain

$$\tau_v = (1-\nu)^{-1} \left[\frac{dt}{(1+t)} - \nu g \left[\left(\frac{(1-\nu)+(1-g)}{1-\nu g} \right) \cdot \left(1 - \frac{\sigma(1-\nu)(1-g)}{(1-\nu)+(1-g)} \right) \right] \frac{dm}{(1+m)} \right] \tag{15}$$

The percent change in value added τ_v differs from τ_π and τ_l only in the second term of the expression and this difference is due to the assumption of a nonunitary elasticity. The term $(1-\nu g)$ is unambiguously positive and less than one; but, the term $[1-\sigma(1-\nu)(1-g)/((1-\nu)+(1-g))]$ may be positive or negative. Thus, increases in m will lead to decreases in value added if the term $[1-\sigma(1-\nu)(1-g)/((1-\nu)+(1-g))]$ is positive; that is if $\sigma < (1-\nu)^{-1} + (1-g)^{-1}$, where $2 \leq (1-\nu)^{-1}$

$+(1-g)^{-1}$ as $0 \leq \nu,\ g \leq 1$. If, therefore, $\sigma < 2$ which includes the Cobb-Douglas function as a special case, increases in the tariff on imported inputs will decrease value added. The opposite result follows if the inequality for σ is reversed.

Comparing τ_q and τ_v, we note that (as in the comparison between τ_π and τ_q) the value of τ_q is less than that of τ_v for positive dt and $dm=0$. The comparison of τ_q and τ_v with respect to the effect of dm is somewhat more complicated. From equation (12) it is clear that $\partial Q/\partial m > 0$, whereas $\partial V/\partial m$ may be either positive or negative. Therefore, depending on the magnitudes of ν, σ, and g, as well as t and m, τ_q may be either less or greater than τ_v. Further, the signs of τ_q and τ_v may differ so that τ_v cannot be used directly even to indicate the direction of change in Q or, a fortiori, indicate the change in the level of imports.

Further, in comparing τ_v and τ_l, similar analysis indicates that the sign of τ_v does not unambiguously indicate the sign of τ_l. We conclude that with nonunitary elasticities of substitution, the traditionally defined effective rate of protection is not even a useful indicator of the signs of the changes in Q and L.

Having considered in detail the continuous forms of the various definitions of effective tariff for the CES case, the discrete forms can be discussed briefly. Using the same procedure as in the Cobb-Douglas case, we obtain

$$(16)\quad \overset{\Delta}{\tau_q} = \left(\frac{1+t}{1+m}\right)^{\lambda} k^{\lambda/(1-\sigma)} - 1,$$

$$(17)\quad \overset{\Delta}{\tau_l} = (1+t)^{\lambda+1}(1+m)^{\sigma-\lambda-1} k^{(1+\lambda-\sigma)/(1-\sigma)} - 1,$$

$$(18)\quad \overset{\Delta}{\tau_\pi} = (1+t)^{\lambda+1}(1+m)^{-\lambda} k^{\lambda/(1-\sigma)} - 1,$$

$$(19)\quad \overset{\Delta}{\tau_v} = (1+t)^{\lambda+1}(1+m)^{-\lambda}\left(\frac{1-\nu g_o k}{1-\nu g_o}\right) k^{\lambda/(1-\sigma)} - 1,$$

where

$$k = \frac{(1+m)^{1-\sigma}}{(1+m)^{1-\sigma} g_o + (1-g_o)},$$

$$g_o = \frac{\delta_1^{\sigma} r_o^{1-\sigma}}{\delta_1^{\sigma} r_o^{1-\sigma} + \delta_2^{\sigma} w^{1-\sigma}} = \frac{r_o X_o}{r_o X_o + w L_o},$$

and

$$\lambda = \frac{\nu}{1-\nu}$$

Note that X_o and L_o in the expression for g_o are the equilibirum free trade values of the imported input and of labor.

Table 2 presents, for six selected values of σ, the tariff rates on the final product (t) necessary to offset the effect of a given tariff on the imported input (m) so as to leave domestc output unchanged. It is interesting, though not unexpected, to find that the necesssary level of t decreases with σ, but the differences in the empirically most relevant range (i.e., $\sigma = 0.5$ to $\sigma = 1.5$) are rather small.

The conclusions concerning the relationships between the various measures of effective tariff are essentially the same as those derived from the analysis of the continuous case. Further, we find that the necessary and sufficient condition for $\overset{\Delta}{\tau_q} \geq 0$ is $(1+t) \geq [g_o(1+m)^{1-\sigma} + 1 - g_o]^{1/(-\sigma)}$. The expression on the right-hand side of the inequality decreases with σ and approaches unity as $\sigma \to \infty$. In other words,

TABLE 2—LEVEL OF TARIFF ON THE FINAL PRODUCT THAT IS NECESSARY TO OFFSET THE EFFECT ON OUTPUT OF A GIVEN TARIFF ON INPUT[a]

Tariff on input	Elasticity of input substitution					
	0.1	0.5	1.0	1.2	1.5	10.0
10%	4.0%	3.9%	3.9%	3.8%	3.8%	2.9%
20%	8.0%	7.8%	7.6%	7.5%	7.3%	4.4%
30%	11.9%	11.5%	11.1%	10.9%	10.6%	5.1%
40%	15.8%	15.2%	14.4%	14.1%	13.6%	5.5%
50%	19.8%	18.8%	17.6%	17.2%	16.5%	5.6%

[a] The entries in the table were calculated on the assumption that $g_o = r_o X_o/(r_o X_o + w L_o) = 0.4$ and that $\nu = 0.9$.

TABLE 3—VALUES OF τ_q^Δ AND τ_v^Δ FOR VARIOUS CHOICES OF ELASTICITY OF INPUT SUBSTITUTION AND NOMINAL TARIFF RATES[a]

t	$\sigma=0.1$		$\sigma=0.5$		$\sigma=1$		$\sigma=1.2$		$\sigma=1.5$		$\sigma=10$	
	τ_q^Δ	τ_v^Δ	τ_q^Δ	τ_v^Δ	τ_q^Δ	τ_v^Δ	τ_q^Δ	τ_v^Δ	τ_q^Δ	τ_v^Δ	τ_q^Δ	τ_v^Δ
0	−0.498	−0.526	−0.491	−0.501	−0.481	−0.481	−0.477	−0.470	−0.471	−0.457	−0.322	−0.051
0.05	−0.221	−0.231	−0.210	−0.195	−0.195	−0.155	−0.188	−0.137	−0.181	−0.114	0.051	0.546
0.10	0.184	0.229	0.201	0.280	0.223	0.345	0.234	0.374	0.245	0.411	0.597	1.462
0.20	1.591	1.935	1.627	2.055	1.676	2.211	1.701	2.281	1.725	2.370	2.496	4.878
0.30	4.324	5.534	4.399	5.801	4.501	6.151	4.550	6.304	4.598	6.499	6.184	12.087
0.40	9.375	12.712	9.521	13.272	9.718	14.005	9.815	14.327	9.912	14.742	12.999	26.463
0.50	18.302	27.055	18.575	27.452	18.941	28.911	19.121	29.553	19.301	30.380	25.045	53.745

[a] The entries in the table were calculated under the assumptions that $g_o = r_oX_o/(r_oX_o+wL_o) = 0.4$, $\nu = 0.9$, and $m = 0.2$.

if the imported input and labor are perfect substitutes, any positive tariff on the final product will afford protection to output regardless of the size of the tariff on the imported input. The same holds true with respect to labor, profit, and value added.

A more detailed comparison between τ_q^Δ and τ_v^Δ is provided in Table 3, which gives the values of these two measures for different values of t and σ. The results indicate that both measures increase with increases in t. However, in most cases τ_v^Δ exaggerates the amount of protection relative to τ_q^Δ. Further, the divergence between the two measures increases with σ (for $t>0$) and with t. These results reinforce our earlier conclusion that the amount of protection afforded to value added is not the most useful measure of effective tariff under conditions of input substitution, and less than perfectly elastic supply of output.

III. Conclusion

The theory of effective protection, as developed and applied in the last decade is formulated around the case of fixed input coefficients, where the output level is determined by short-run supply constraints on the inputs. For that case, and that case alone, the protection on value added is a correct indicator of the economic cost of the tariff. Attempts to "adapt" the definition to other types of production functions (see, e.g., James Anderson and Seiji Naya; Leith (1968, 1971)) cannot succeed unless one starts from the fundamental decision-making process of the firm.

In line with received economic theory, this paper uses a profit maximizing model for the firm to allow for simultaneous changes in output price and marginal cost which reflect the changes in the tariffs on both the output and the input. With particular reference to the Cobb-Douglas and *CES* production functions, the paper develops mathematical expressions for the changes in domestic production, and in the level of labor input. These results are then used to formulate and compare alternative definitions of effective protection (labeled τ), i.e.: the proportional change in value added (V), domestic output (Q), labor input (L), and profits (Π). For each variable both continous and discrete cases are considered

Specific expressions for these definitions of effective tariffs are developed for both the Cobb-Douglas and the *CES* production functions. In the Cobb-Douglas case it is shown that the protection accorded to value added exaggerates the effect of the tariff structure on domestic output and therefore on resource allocation, although it is a good indicator of the protection accorded to labor. In the case of the *CES* production function, the protection accorded to value added does not even indicate the direction of change in domestic production, the level of imports or the level of labor input. The expressions developed in this paper are, at least in the context of our profit maximizing model, better indicators of these changes.

REFERENCES

J. Anderson and S. Naya, "Substitution and Effective Protection," *Amer. Econ. Rev.*, Sept. 1969, *59*, 607–11.

B. Balassa, "Tariff Protection in Industrial Countries: An Evaluation," *J. Polit. Econ.*, Dec. 1965, *73*, 573–94.

———, **S. E. Guisinger, and D. M. Schydlowski**, "The Effective Rate of Protection and the Question of Labor Protection in the United States: A Comment," *J. Polit. Econ.*, Sept./Oct. 1970, *78*, 1150–63.

G. Basevi, "The United States Tariff Structure: Estimates of Effective Rates of Protection of U.S. Industries and Industrial Labor," *Rev. Econ. Statist.*, May 1966, *68*, 147–60.

W. M Corden, "The Structure of a Tariff System and the Effective Protective Rate," *J. Polit. Econ.*, June 1966, *74*, 221–37.

H. G. Johnson, "The Theory of Effective Protection and Preferences," *Economica*, May 1969, *36*, 119–38.

J. C. Leith, "Substitution and Supply Elasticities in Calculating the Effective Protection Rate," *Quart. J. Econ.* Nov. 1968, *82*, 488–601.

———, "The Effect of Tariffs on Production, Consumption, and Trade: A Revised Analysis," *Amer. Econ. Rev.*, Mar. 1971, *61*, 74–81.

——— **and G. L. Reuber**, "The Impact of the Industrial Countries' Tariff Structure on Their Imports of Manufactures from the Less-Developed Areas: A Comment," *Economica*, Feb. 1969, *36*, 75–80.

Deposit Expansion and Federal Reserve-Banking System Interaction: A Micro Unit Simulation

By WILLARD T. CARLETON AND WILLIAM R. BRYAN*

Over the past several years, a prodigious research effort has created several operational models depicting significant features of the monetary-economic process. The FRB-MIT (see Frank de Leeuw and Edward Gramlich) and Federal Reserve Bank of St. Louis (see Leonall Andersen and Keith Carlson) models come most readily to mind, though there are others. Now that these models are on line, the research effort has shifted in two directions. On the one hand, attention has turned toward simulation experiments as a means of learning about the properties of the models and as a means of tracing out the implications of alternative monetary and fiscal policies (see George Kaufman and R. D. Laurent). On the other hand, work is moving forward to enrich subsystems of the large scale models (see Franco Modigliani, Robert Rasche, and Jerome Cooper).

This paper proposes a modest departure from the research that has gone into and appears to be coming from these models of the monetary-economic process. Despite our concurrence with the spirit of this recent research, we would argue that the work poses several problems; first it is conducted exclusively at an aggregative level, and second it assumes Federal Reserve policy actions to be exogenous. In an earlier article (Bryan and Carleton) we suggested:

1) The micro bank response is exactly inverse to the macro bank response to monetary policy. That is, the major disequilibrating force felt by an individual bank is a change in deposits; the adjustment mechanism involves changes in its reserve position and its holdings of earning assets. At the aggregate level, however, the Federal Reserve System directly alters the magnitude of member bank reserves and aggregate response takes the form of changes in total deposits and earning assets.

2) In a long-run equilibrium model, this distinction is of little empirical importance; one assumes a completed banking system response to monetary policy. Accordingly, the usual macro specification contains deposits as the dependent variable and reserves as the independent variable.[1] However, the research cited above—in which bank behavior has been hypothesized explicitly—appears to have been motivated by concern with the dynamics of the monetary mechanism, with

---

* Professor of finance and economics at the Amos Tuck School of Business Administration, Dartmouth College, and professor of finance, University of Illinois, currently on leave with the Treasury Department, respectively. The order in which our names are given implies nothing as to chronological seniority or relative responsibility for the present article. We are indebted to the Tuck School Associates Program for financial support and to James Guerber, Dominique Lautier, and Jay Cary for computer programming.

[1] It is interesting to note that the studies by Karl Brunner and Allen Meltzer, while professing an interest in monetary dynamics, not only postulate a macro bank model, but estimate relationships from annual data. Similarly, the Modigliani-Rasche-Cooper study professes a concern with the short-run adjustments of an individual bank, but then makes estimates with aggregate quarterly data.

such issues as lags in the effects of monetary policy. In such work, slippages caused by differential bank behavior would appear to be of the essence, but before such slippages can be captured the specification must be reversed. If the short-run bank deposit-credit expansion process is of interest to monetary theorists, then its articulation and testing at some level other than purely aggregative would appear necessary.

But perhaps the ultimate misconstruction results from the general failure of the profession to take explicit account of the interaction between monetary actions and banking system responses.[2] We continue to wallow in the longstanding and widespread habit of assuming Federal Reserve actions to be exogenous—actions manifested in changes in the monetary base, total reserves, unborrowed reserves, Federal Reserve credit, etc. Such practices enmesh us in simultaneous equations bias and related difficulties. The fact that the Federal Open Market Committee (*FOMC*) could, in principle, require the manager of the Open Market Account to buy or sell $\$X$ million of securities before the next Committee meeting is not a sufficient excuse for assuming such is observed practice. So long as the *FOMC*'s directive remains imprecise, its control over the Account Manager imperfect, and its concern with "defensive" operations manifest, then econometric models err badly when they assume as "fixed" observed changes in bank reserves, or (alternatively) when they assume zero covariance between such changes and any recognized jointly dependent variable.

The prescription arising from these criticisms is imposing. It would appear that the minimum adequate model for studying the central and commercial banking system parts of the monetary mechanism would:

1. treat Federal Reserve and bank behavior as output of an interdependent system;
2. concern itself with the dynamics of short-run behavior, probably producing bank deposit expansion/contraction as disequilibrium phenomena;
3. specify the banking system portion of the model in terms of micro unit balance sheet decisions together with market-clearing mechanisms.[3]

Data limitations currently preclude complete realization of our monetary's researcher's *summum bonum*; a belated and sterilized version of relevant Federal Reserve actions is a matter of public record, but weekly statements of condition of commercial banks are treated as confidential. Our earlier work suggested that micro unit bank models would be fruitful for monetary research. In the present paper we have had to abandon a purely empirical approach. In what follows, we have combined an empirical model of Federal Reserve open market purchases and sales with an arbitrarily specified micro unit model of the commercial banking system in order to simulate the behavior of jointly produced variables. The results are both reasonable and suggestive of the benefits to be generated by disaggregated approaches to studying the origin and transmission of monetary actions.

The simulation presented below involves a scheme for aggregating, or "adding up," underlying micro behavior. At the level of the individual commercial bank we "explain" changes in its holdings of earning assets, excess reserves, and in its total borrowing. For the banking system we simulate the deposit expansion-and-contraction process along with aggregate borrowing from the Federal Reserve. In addition, we depict the operation of the Federal funds market, thereby generating a Federal funds rate. Finally, our overall model is made a closed system by treating Federal Reserve open market operations as endogenous.

---

[2] This is not to say that nobody understands this interaction. A. James Meigs, Jack Guttentag, and John Wood have developed this theme fully. Unfortunately, their voices get lost in the wilderness, and their admonitions do not get adequately imbedded in large-scale models

[3] The Modigliani-Rasche-Cooper study moves in this direction. However, the authors assume that aggregate changes in unborrowed reserves are exogenous; they then regard free reserves and interest rates as endogenous. Had the Federal Reserve been acting on free reserves—as some argue that they were, at least in the very short run (e.g., Meigs)—then unborrowed reserves would have been endogenous.

The structural equations of the model are specified in every detail. As in other studies, we select the dependent and independent variables and specify the form of the equations. Unlike other studies, however, we also specify most of the values of the parameters of the system. That is, aside from our prediction equation for Federal Reserve open market operations, we do not estimate parameter values by the usual econometric methods.

One factor which undermines confidence in simulation models of this sort is the feeling that the researchers simply fudge the model until it works. Our model cannot fully escape this criticism. However, we faithfully observed two criteria in our specifications. First, it was necessary for our underlying micro behavior to be both sensible and well within the body of received doctrine. Thus, there are no novel or special purpose explanations of the behavior of individual banks.

Secondly, it was necessary for our aggregating mechanism to preserve the integrity of distinctions between endogenous and exogenous variables. There are variables—e.g., changes in demand deposits—which are generally thought to be exogenous to the individual bank but which in the aggregate are regarded as endogenous to the banking system. Conversely, there are variables—e.g., a bank's holdings of total reserves—which are generally thought to be endogenous to the individual bank but which in the aggregate are regarded as exogenous to the banking system. Our simulation takes care to violate none of these considerations.

One might wonder why in selecting parameter values we did not take explicit advantage of the results of the several banking models that have been specified and tested over the past few years. After all, the behavioral statements undergirding these models have had their origin at the micro level. That is, intuition or insights regarding the behavior of individual banks have guided the process of formulating hypotheses. Unfortunately, the testing of these models has been carried out with data generated by aggregates. As we suggested above, there are frequent reversals of causation in the aggregation process. Unless we are prepared to believe that all observations depict equilibirum conditions, there are substantial aggregation problems. We are left believing that the available parameter estimates have questionable value for our purposes.

I. The Structure of Federal Reserve Decisions

Description of Federal Reserve open market purchases and sales as an endogenous variable is a trial and error process for at least two reasons. The first is conceptual. Perhaps blinded by the textbook simplicity of models taking the quantity of money (or its rate of change, or more recently the monetary base) as given, theorists have rarely attempted to formalize the implicit model(s) contained in the historical record of monetary policy actions.[4] In addition, informed explanations —particularly those of Federal Reserve spokesmen—emphasize the multidimensional and poorly ordered preferences of policy makers, as well as the presence of a number of constraints on their decisions.[5] Consequently, we are left with little prior work to guide us. Secondly, and especially in view of the informality of available descriptions of the Federal Open Market Committee–Open Market Account behavioral nexus, one may appeal to a variety of often contradictory sources for information on the structure of Federal Reserve decisions.

---

[4] For an excellent exception, see Wood.

For the limited demonstration purposes of this study we have assumed that discount rate and reserve requirement changes were determined outside the system during the period under review. The period of monetary behavior examined, January 1967 through mid-June 1968, was deliberately chosen to include what Federal Reserve spokesman would (and did) consider episodes of active "ease" and "tightness." The period encompasses changes in the discount rate and reserve requirements. It avoids recent institutional developments regarding the Euro-dollar market. Following the interpretation by the Federal Reserve Bank of St. Louis (see Leonall Andersen and Michael Rigg, Jerry Jordan and Charlotte Ruebling) an interpretation based on *FOMC* directives, banking and interest rate statistics, and knowledge of *FOMC* members' qualitative assessments—the weekly record of Fed actions was broken into four sub-periods to represent four different overall policies:

Period—Weeks Ending	Policy Consensus
1) Jan. 10, 1967–May 2, 1967	Ease
2) May 2, 1967–Aug. 15, 1967	Maintain conditions
3) Aug. 15, 1967–Nov. 14, 1967	Maintain conditions, but avoid bank credit expansion
4) Nov. 14, 1967–June 18, 1968	Maintain and/or firm up money market conditions

Open market operations result from a complex and intuitive decision-making process. We make no effort here to replicate that process. However, we do seek to represent the notion that changes in Federal Reserve holdings of government securities emerge, at least in part, from a milieu of operating predilections, dynamic policy goals, general economic conditions, and short-run, transitory "shocks." Our choice of explanatory variables (described in Table 1) was designed to incorporate proxies for these factors. Multiple regressions were used to explain weekly changes in daily average Federal Reserve holdings of *U.S.* Government securities for each of the four periods. Our choice of a linear statistical model and its required assumptions is at best only an approximation of this process.

It is our view that a major operating predilection of the Federal Reserve is that it tends to resist abrupt money market changes. In light of this we expect open market operations to be geared toward offsetting factors that might serve to induce sharp changes in interest rates. Such considerations led us to expect free reserves, FR_t, to be negatively related to the dependent variable, OMP_t (open market purchases and sales). These expectations were based on the notion that as market interest rates begin moving up, causing FR_t to go negative, the Federal Reserve will purchase Government securities to moderate the decline in free reserves and the rise in interest rates; conversely, as credit conditions weaken, the Federal Reserve will sell to moderate increases in free reserves and declines in interest rates.[6]

Similar considerations led us to expect OMP_t to be positively related to the differ-

---

[5] See *The Federal Reserve and the Treasury: Answers to Questions from the Commission on Money and Credit.*

[6] For a discussion of this strategy, see Guttentag.

TABLE 1—DESCRIPTION OF REGRESSION VARIABLES

Symbol	Description
OMP_t	Change in daily average Federal Reserve holdings of *U.S.* Governments, resulting from open market purchases and sales. Sources and uses of reserves are summarized by the identity $\Delta TR_t - OMP_t - \Delta RB_t = MF_t$ where ΔTR_t = change in total member bank reserves, ΔRB_t = change in member bank borrowing, and MF_t = market factors, the net of all other sources and uses of reserves.
$RDIF_t$	Difference between daily average 3-month Treasury bill rate and the maximum of Fed discount rate (New York) and Federal funds rate, both rates lagged one week. $RDIF_t$ is thus an opportunity cost for banks under reserve pressure.
FR_t	Free reserves for week $t-1$, by market lore and sometime Federal Reserve interpretation a measure of money market conditions.
$TRGR_t$	$(TR_{t-1} - TR_{t-3})/TR_{t-3}$. This lagged two week percentage change in total reserves is designed to reflect the longer-run setting within which current open market operations are occurring.
IP_t	Industrial production index, lagged one month till the last week of the month, at which time the current month's index is used.
MF_t	Market factors affecting reserves during current week. This variable is not lagged because, while it is clearly exogenous within one week periods, open market forecasts and practices are directed toward contemporaneously offsetting this variable as a market-disruptive influence.

Source: Federal Reserve Bulletin

ence between market interest rates and the cost of borrowing, $RDIF_t$, during periods of ease. However, during periods of severe credit restraint we might expect the relation between $RDIF_t$ and OMP_t to be reversed. For, during such periods the federal funds rate is likely to rise above yields on Treasury bills. Moreover, it is likely that changes in credit demands will be substantially mirrored in movements in this rate. If under such circumstances the Fed continues to moderate movements in interest rates the coefficient relating $RDIF_t$ to OMP_t will take on a negative sign.

Results with FR_t were consistent with hypotheses and generally significant. Results with $RDIF_t$ were somewhat disappointing though they were generally consistent with our hypothesis regarding the expected coefficient signs—even to the extent of showing a negative sign in period 3). However, the coefficients did not quite reach acceptable levels of significance.

We expect a positive relation between the percentage change in total reserves over the preceding two weeks, $TRGR_t$ and the dependent variable. $TRGR_t$ is intended to depict the monetary policy setting. On the average we expect week-to-week open market operations to move in the same direction as underlying changes in total reserves. Indeed, over reasonably long periods we expect OMP_t to be the principal factor bringing about changes in reserves. Results were consistent with hypotheses and generally significant.

There is a long standing notion that monetary policy tends to "lean against the wind." If this is the case, we should expect a negative relation between general measures of business activity such as the industrial production index, IP_t, and open market operations. Results were generally consistent with hypothesis. However, in only one test did the coefficient reach an acceptable level of significance.

As indicated, we expect the Federal Reserve to offset short-run shocks to the reserve base. Hence, we expect a negative sign to attach to the coefficient relating

TABLE 2—REGRESSIONS EXPLAINING CHANGES IN FED HOLDINGS OF *U.S.* GOVERNMENT SECURITIES, OMP_t

Variable	Period 1)	2)	3)	4)
Intercept	16232.8 (2.16)	11180.8 (.41)	−790.7 (.11)	9724.3 (1.02)
$RDIF_t$	27405.6 (1.05)	4434.5 (.22)	−49104.3 (1.53)	22570.8 (1.18)
FR_t	−.622 (1.54)	−1.02 (1.87)	−1.44 (2.91)	−.766 (2.45)
$TRGR_t$	6027.5 (1.97)	6823.6 (1.56)	8164.6 (2.54)	6094.7 (2.44)
IP_t	−102.0 (2.13)	−69.9 (.40)	8.66 (.19)	−60.6 (1.03)
MF_t	−.478 (4.00)	−.752 (4.94)	−.910 (7.80)	−.439 (2.45)
R^2	.73	.72	.87	.36
degrees of freedom	10	9	9	23

Note: Absolute values of *t*-ratios are given beneath the coefficient estimates in parentheses.

open market operations to market factors affecting reserves, MF_t. Results were consistent with hypotheses and statistically significant.

Results from this regression work are used below in a model in which Federal Reserve purchases and sales of Government securities turn out to be endogenous to the total "system" and in which there is interaction between actions of the Fed and the behavior of the banks. In the simulation the regression coefficients shown in Table 2, together with successive weekly values of the independent variables, are used to generate weekly purchases and sales of Government securities. Because several of the independent variables—namely FR_t, $RDIF_t$, and $TRGR_t$—are generated, at least in part, by the banking system portion of our model, open market operations may be regarded as endogenous. In turn, because open market operations affect the credit expansion process there is feedback and interaction between the nominal dependent variable and the so-called independent variables.

Notwithstanding our views regarding interaction and feedback, ordinary least squares procedures were used to estimate the structural coefficients. In light of our expressed concern regarding work that assumes Federal Reserve behavior as exogenous and then goes on to estimate, say, a banking system response, perhaps we should express misgivings at our own empirical work. In principle, we are no less subject to criticism in treating the rest of the world as exogenous in estimating a Federal Reserve behavioral equation. We have sought to turn aside some of our self-criticism by limiting ourselves to lagged independent variables. In any event, we make no pretense that our model specification is satisfactory for other than the demonstration purposes of this study.

II. The Structure of Individual Bank Decisions

The first thing needed for a micro unit decision model is a balance sheet. The degree of balance sheet aggregation for this model was dictated by two considerations: first, an overriding interest in those central and commercial banking system interactions which produce demand deposit expansion and contraction, rather than in the details of bank earning asset allocations; and second, a felt need to avoid any disaggregation not absolutely required in order to serve the first objective. With more balance sheet detail would come the requirement for more decision rules, and ours were assumed, not empirically estimated.

Balance Sheet of Individual Bank

Assets

- Cash items in process of collection (*CIPC*)
- Reserves
  - Required (RR)
  - Excess (RX)

or
  Unborrowed (RU)
  Borrowed (RB)
Federal funds sold (FF^-)
Earning assets (EA)

Liabilities

Demand deposits (DD)
Time certificates of deposit (CD)
Other time and savings deposits (SD)
Federal funds purchased (FF^+)
Borrowing from the Fed (RB)
Other items (OI)

We depict our typical bank as receiving, exogenously, changes each day in its DD, CD, SD, and $CIPC$ positions. Given reserve requirements and stability in OI (which primarily consists of capital less fixed assets), the bank thus faces a daily funds available (ΔF) equation:

$$\begin{aligned} \Delta F_{it} &= \Delta DD_{it} + \Delta CD_{it} + \Delta SD_{it} \\ &\quad - \Delta CIPC_{it} - \Delta \overset{*}{R}X_{it} \end{aligned} \tag{1}$$

where Δ is a backwards first difference operator and $\Delta \overset{*}{R}X_{it}$ is the change in desired excess reserves from day t−1 to day t.

The bank will make a subsequent purchase or sale of earning assets (according to rules to be described below), generating an implied desired Federal funds purchase or sale for the day:

$$\Delta EA_{it} - \Delta F_{it} = \overset{*+}{FF}_{it} - \overset{*-}{FF}_{it} \tag{2}$$

It should be recognized that we specify this sequence arbitrarily. One could imagine circumstances in which it would be appropriate to reorder the events. For example, it seems plausible to imagine that—in light of high returns for lending Federal funds—a bank might reduce its holdings of other earning assets. Notwithstanding such considerations, we specify that if ΔEA_{it} exceeds ΔF_{it} then $\overset{*+}{FF}_{it}$ will be positive for the day and $\overset{*-}{FF}_{it}$ zero. Conversely if ΔEA_{it} is less than ΔF_{it} then $\overset{*+}{FF}_{it}$ will equal zero, and $-\overset{*-}{FF}_{it}$ will equal the discrepancy. Several features of this process should be noted:

1) At the level of the individual bank we make no distinction between sources of borrowed reserves. It therefore is convenient to summarize any day's implied desired borrowing in $\overset{*+}{FF}_{it}$. However, as will be discussed below, we determine borrowing from the Fed in our aggregation procedure.

2) Our market-clearing mechanism does not *formally* operate except at the close of each "statement" week, on Wednesday. This has two consequences. First, $\overset{*+}{FF}_{it} - \overset{*-}{FF}_{it}$ could be regarded as the day t's real spot reserve excess or deficiency—to be covered by an appropriate transaction at some time within the reserve settlement period. Secondly, because Federal funds are not *formally* purchased or sold each day, the funds statement, equation (1), does not contain any unwinding of the previous day's Fed funds transactions. As will be seen, final settlement does take place on Wednesday, and the daily funds equation for each bank on the next day will incorporate such unwinding.

3) Conceptually, we do not preserve derivative deposits at the level of the individual bank. That is, all proceeds from purchase or sale of earning assets during day t are assumed to be cleared the same day (if you will, settlements are made in Federal funds). The output to the rest of the banking system of this deposit write-up/write-down and shift to another bank will be accounted for in our aggregating functions.

From the above description it can be seen that two decisions, $\Delta \overset{*}{R}X_{it}$ and ΔEA_{it}, summarize each bank's daily response to changes in its operating environment. Before specifying the functions that generate these decisions it is necessary to describe the simulation model's population of banking units. For programming convenience, we assumed a common weekly reserve settlement period for all banks (a departure from Federal Reserve operating rules during the period of the simulation). However, in deference to the popular notion that distributional effects may emerge by virtue of differences among banks, we created two classes of banks—large, money center units and small, conservatively run units. Thus our final model includes twenty-five of each kind. Initial balance sheet proportions for the large banks were modelled after the largest twenty-five banks as of December 31, 1966. The smaller units were based on the relative frequency of nonreserve city banks grouped by size of total deposits.

These units were scaled up to represent the approximate size of the banking system, including member and nonmember banks, as of year-end 1966. Initial Federal funds, borrowing from the Federal Reserve and excess reserves were allocated in the following fashion: Federal funds positions were assumed to be zero for all banks. This assumption no doubt produced some of the divergencies between observed and simulated borrowing from the Federal Reserve in the first few weeks of the simulation runs. Initially, this borrowing was prorated to large banks and excess reserves were prorated to small banks. This was felt to be a reasonable approximation to observed bank behavior.

Decision functions and constraints differed between each class of bank. For the large banks $\Delta \overset{*}{RX}_{it}=0$, while for the small banks ΔRX_{it} was treated as a random variable with uniform distribution, and $.2RX_{it-1}<\Delta \overset{*}{RX}_{it}<1.8RX_{it-1}$. This latter specification might not be adequate in a longer-run model in which one can observe a systematic decline in excess reserves, reflecting simultaneously a rise in interest rates and an increase of funds management efficiency in the banking system. For a daily model running over a year or so, however, we think the assumption that aggregate ΔRX is random is a very close approximation to reality.

Earning asset decisions are more complex, since it is desirable to incorporate both funds availability and opportunity cost considerations. The decision functions for earning assets were specified to be the same for small and large banks, and of the form

$$(3)\quad \Delta EA_{it} = \delta_{it}\Delta F_{it}$$

where

$$(4)\quad \begin{aligned}\delta_{it} &= -15\,[r_{it} - \text{maximum}\,(r_{2t}, r_{3t})]\\ &\quad + e_{1it} \quad \text{for } \Delta F_{it} < 0;\end{aligned}$$

and

$$(5)\quad \begin{aligned}\delta_{it} &= 15\,[r_{1t} - \text{maximum}\,(r_{2t}, r_{3t})]\\ &\quad + e_{2it} \quad \text{for } \Delta F_{it} \geq 0;\end{aligned}$$

and

r_{1t} = daily average three-month Treasury bill rate for the current reserve settlement week;

r_{2t} = New York Federal Reserve Bank discount rate for the current reserve settlement week (changes in this rate were assumed to occur on the first day of the relevant week);

r_{3t} = Federal fund rate for the current day (this rate is generated daily, within the model, after the initial day of the simulation);

e_{1it} = random variable, uniformly distributed with $.4<e_{1it}<1.4$;

e_{2it} = random variable, uniformly distributed with $.5<e_{2it}<1.5$.

It is clear from this description of earning asset decisions and from equation (2) that cumulative under- or overreaction to funds flows can produce a large cumulative desired Federal funds position. Since we know that the ground rules under which bank money desk managers typically operate set limits on exposure to the risks inherent in large reserve position shifts, we placed the following constraints possible changes in purchase or sale of Federal funds.

$$
\begin{aligned}
& -.15(RR_{it-1} + \overset{*}{RX}_{it-1}) \\
(6) \qquad & \leq \sum_{\gamma=1}^{t} (\overset{*}{FF}_{i\gamma} - \overset{*}{FF}{}^{-}_{i\gamma}) \\
& \leq .1(RR_{it-1} + \overset{*}{RX}_{it-1})
\end{aligned}
$$

where t = 1, . . ., 7.

That is, cumulative net desired lending will not exceed 15 percent of the previous day's spot reserve position, nor cumulative net desired net borrowing exceed 10 percent of the previous day's reserve position. Daily earning asset purchases and sales as described in equation (2) are adjusted, when necessary, so as to satisfy constraint (6).

Verbally, the model of banking behavior depicted with equations (3), (4), and (5) and inequalities (6) is of a bank that buys and sells earning assets in response to daily funds inflows and outflows, respectively. The degree of this response, as given by the δ coefficients, would be randomly distributed around a mean of 1.0 in the absence of opportunity cost considerations. As the opportunity cost (the spread between asset opportunities, r_{it}, and potential borrowing costs, r_{2t} and r_{3t}) becomes positive the bank's behavior becomes more aggressive. Conversely, a negative spread induces more conservative behavior. For example, with a positive spread of .005 (which was about the largest value in the period under study), the bank would be more willing to run the risks of incurring a reserve deficit and would tend to overbuy (δ averaging 1.075 in equation (5)) and undersell (δ averaging .925 in equation (4)) earning assets.

While the above specification captures much of the flavor of the real world, it must be emphasized that it has no empirical basis other than current lore of banking and conversations with practicing bankers. An econometric model of an individual bank would no doubt be richer in detail and would be specified to change over time. Banks probably are infinitely elastic suppliers of commercial loans to their customers on existing terms *for short periods of time*. Over a month or a quarter or six months, however, one would expect to see some adjustment, resulting from changes in bank expectations and customer demands for loans, in equations such as (4) and (5) which include lending behavior implicitly. Some of the discrepancies between actual and simulated banking system behavior reported in Section IV below probably are a consequence of our specification of bank decision rules as unchanging over the simulation period.

III. The Structure of Federal Reserve-Banking System Interaction

As stated in Section II, at the level of the individual bank we take $\Delta CIPC_{it}$, ΔDD_{it}, ΔSD_{it} and ΔCD_{it} as exogenous. This means that the overall model must provide a mechanism for allocating these variables among the banks. The equations which are used to generate these variables are as follows:

$$(7) \quad \Delta CIPC_{it} = e_{3it} DD_{it} - CIPC_{it-1}$$

where e_{3it} is a random variable drawn from a uniform distribution with $.09 < e_{3it} < .11$. Most empirical studies of banking system behavior ignore *CIPC* by netting it from demand deposits. For very short-run and disaggregated models realism dictates that we include this important source of variability in available funds.

$$(8) \quad \Delta DD_{it} = \left\{ \frac{DD_{it-1}}{\sum_{i=1}^{50} DD_{it-1}} \right\} \alpha_1 + e_{4it}$$

$$(9) \qquad \alpha_1 = 1/7\,(OMP_t + MF_t) + \sum_{i=1}^{50} \Delta EA_{it-1} - \sum_{i=1}^{50} e_{4it}$$

where e_{4it} is normally distributed with $\mu = 0$, $\sigma^2 = .005DD_{it-1}$ for small banks, $\sigma^2 = .01DD_{it-1}$ for large banks. Equations (8) and (9) reflect two key modelling assumptions: that bank i's demand deposit experience is at least roughly proportionate to its size; and, as mentioned previously, that the derivative deposits created by all banks on day $t-1$ are immediately drawn down to be redistributed over the system on day t. Thus we can think of day t's aggregate demand deposit change as made of two components, the first one exogenous to the banking system, the second reflecting the initial round of endogenous expansion or contraction arising from day $t-1$'s bank credit decisions, i.e., the first round of the textbook model at work.

$$(10) \qquad \Delta SD_{it} = \left[\frac{SD_{it-1}}{\sum_{i=1}^{50} SD_{it-1}} \right] \Delta SD_t$$

where ΔSD_t is the aggregate change in savings and other non-*CD* time deposits, assumed to be exogenous for the week in which day t occurs. The observed stability of these deposits (except perhaps around quarterly interest payment dates) led us to dispense with random changes.

For time certificates of deposit (*CD*), we assumed the following:

$$(11) \qquad \Delta CD_{it} = \left[\frac{CD_{it-1}}{\sum_{i=1}^{25} CD_{it-1}} \right] \alpha_2 + e_{5it}$$

$$(12) \qquad \alpha_2 = \Delta CD_t - \sum_{i=1}^{25} e_{5it}$$

where ΔCD_t is the aggregate change in daily average time certificates of deposit for the week in which day t occurs. Our treatment of ΔCD_t as exogenous is a necessary compromise with reality, since an adequate framework for endogenous determination of such deposits would have to incorporate markets for short-term debt instruments more generally. The variate e_{5it} is normally distributed with $\mu = 0$ and $\sigma^2 = .01CD_{it-1}$. Only the large banks in our model population are given a *CD* position.

The interest rate on Federal funds, while taken by each bank as given, is changed daily as an endogenous variable of the system. In designing this part of the model we faced two choices: to portray the rate for each day as a market clearing rate, with the implied assumptions of equilibrium and well-defined aggregate demand and supply functions for Federal funds. Given the incompatibility of such a framework with our description of daily micro level bank decisions in a *disequilibrium* context, however, we chose instead to formulate the Federal funds market as one in which the rate—or price—adjusts with a one-day lag to the discrepancy between aggregate quantities supplied and demanded. The following rules summarize our (admittedly *ad hoc*) procedures:

1) If $\sum_{i=1}^{50} (\overset{*}{FF}{}^{+}_{it-1} - \overset{*}{FF}{}^{-}_{it-1}) \geq 0$ and

$r_{1t} \geq \text{maximum } (r_{2t}, r_{3t-1})$, then

$r_{3t} = r_{3t-1} + .0003$

2) If $\sum_{i=1}^{50} (\overset{*}{FF}{}^{+}_{it-1} - \overset{*}{FF}{}^{-}_{it-1}) \geq 0$ and

$r_{1t} < \text{maximum } (r_{2t}, r_{3t-1})$, then

$r_{3t} = r_{3t-1} + .0001$

3) If $\sum_{i=1}^{50} (\overset{*}{FF}{}^{+}_{it-1} - \overset{*}{FF}{}^{-}_{it-1}) < 0$ and

$r_{1t} \geq \text{maximum } (r_{2t}, r_{3t-1})$, then

$r_{3t} = r_{3t-1}$

4) If $\sum_{i=1}^{50} (\overset{*+}{FF}_{it-1} - \overset{*-}{FF}_{it-1}) < 0$ and

$r_{1t} <$ maximum (r_{2t}, r_{3t-1}), then

$r_{3t} = r_{3t-1} - .00025$

5) $r_{3t} \geq .025$

Since the Federal funds market is not formally cleared until the last day of each week, one can think of rules 1) through 5) as summarizing interior recontracting. Moreover, given decision rules summarized by equations (3), (4), and (5), adjustment of r_3 from day t−1 to day t tends to produce the kind of earning asset choices one would expect from such changes in opportunity costs. For example, if the Treasury bill rate, r_{1t}, is consistently higher than the discount rate, r_{2t}, or Federal funds rate, r_{3t}, one would expect (and the model generates) an aggregate tendency of the banking system to overbuy earning assets. The residual net demands this behavior produces (bank credit and demand deposits increasing more rapidly than can be validated by observed increases in the stock of unborrowed reserves) will force increases in the Federal funds rate. Our statement of rules 1)–5), and equations (3)–(5) in terms of the maximum of r_{2t} and r_{3t} reflects a conservative—and nonempirical—specification of discontinuities in opportunity cost considerations.

It is clear that on any given day t the banking system's net demand for Federal funds can be positive or negative. Moreover, the existence of such a net demand/supply has implications for individual banks, the system as a whole, and the Federal Reserve. Consider, for example, the case in which $\sum_{i=1}^{50} (\overset{*+}{FF}_{it} - \overset{*-}{FF}_{it}) > 0$. In such a case both borrowers and lenders could be satisfied; lenders for the obvious reason and borrowers because there exists a last resort source of reserve balances, the Federal reserve discount window. Thus, the amounts $\overset{*+}{FF}_{it}$ and $\overset{*-}{FF}_{it}$ are added to each bank i's cumulative desired Federal funds position, $\sum_{\gamma=1}^{t} (\overset{*+}{FF}_{t\gamma} - \overset{*-}{FF}_{i\gamma})$.

On the other hand, if $\sum_{i=1}^{50} (\overset{*+}{FF}_{it} - \overset{*-}{FF}_{it}) < 0$, all borrowers can be satisfied, but for some lenders there will be no market, since our "tatonnement" described above is not required to converge to an equilibrium. While it is possible for disappointed lenders to carry their unintended excesses of reserves to day t+1 (the reserve settlement rules cumulate positions within the week), it is also possible to freeze frustrated Federal funds loans forever out of the market by adding them to excess reserves for the day. This is accomplished by entering randomly the list of banks for which $\overset{*+}{FF}_{it} - \overset{*-}{FF}_{it} < 0$ and adding the absolute amount to $\Delta \overset{*}{RX}_{it}$ until in the aggregate $\sum_{i=1}^{50} (\overset{*+}{FF}_{it} - \overset{*-}{FF}_{it}) = 0$. We adopted this latter procedure in order to induce some interest rate response in aggregate excess reserves for the first six days of each week. Market-clearing arithmetic would otherwise call for such treatment on the settlement day. That is, from each bank's daily balance sheets (and assuming that desired RB is held to zero within the week),

$$\Delta RR_{it} + \Delta \overset{*}{RX}_{it} + \overset{*-}{FF}_{it} = \Delta RU_{it} + \overset{*+}{FF}_{it} \tag{13}$$

Rearranging terms and summing over all banks and days,

$$\sum_{t} (\overset{*+}{FF}_{t} - \overset{*-}{FF}_{t}) = \sum_{t} \Delta RR_{t} - \sum_{t} \Delta RU_{t} + \sum_{t} \Delta \overset{*}{RX}_{t} \tag{14}$$

where $\overset{*-}{FF}$ and $\overset{*-}{FF}$ are day t's aggregate desired Federal funds borrowing and lending, respectively. But we also know that on the books of the Federal Reserve

$$RR_t + RX_t = RU_t + RB_t. \tag{15}$$

or

$$\sum_t \Delta RB_t = \sum_t \Delta RR_t - \sum_t \Delta RU_t + \sum_t \Delta RX_t \tag{16}$$

Therefore, if $\sum_t \Delta \overset{*}{R}X_t = \sum_t \Delta RX_t$, or desired excess reserves equal actual excess reserves, we can subtract (16) from (14) to get

$$\sum_t (\overset{*+}{FF_t} - \overset{*-}{FF_t}) = \sum_t \Delta RB_t \tag{17}$$

Thus aggregate net demand for Federal funds, which has a zero lower bound because of our treatment of any day's net negative amounts, equals borrowing from the Federal Reserve. That is, Federal funds transactions cannot create or destroy reserves for a statement week, even though net demands and accompanying bank credit and demand deposits can exist on any given day.

Several features of our market-clearance should be noted:

First, changes in borrowing from the Federal Reserve, $\sum \Delta RR_t$, and hence $\sum RB_t$, are determined without reference to individual banks. Our individual bank decision functions provide for Federal funds borrowing only. If Federal funds lending is inadequate, some banks end up with the "hot potato"; they in fact borrow from the Federal Reserve. This borrowing occurs at the end of the statement week. On the first day of the following week each bank's funds equation (1) has a term on the right-hand side of the form $-\sum_{\gamma=1}^{7}(\overset{*+}{FF}_{i\gamma}-\overset{*-}{FF}_{i\gamma})$, reflecting reversal of the previous day's completed Federal funds transactions (as indicated, borrowing from the Federal Reserve is imbedded in these transactions).

Second, market clearing thus takes place on day seven of each week (other than placing the zero lower bound on aggregate ΔRB_t daily) in the sense that only on this day are transactions recorded which formally require reversal on the next day.

Third, money market lore suggests that there is considerable intra-weekly fluctuation in "pressure" variables (Federal funds rates, bank reserve positions, etc.). Weekly clearing allows this to be reproduced in the model; daily clearing would involve daily repayment of transactions involving reserves, and restrict the fairly substantial daily movements observed in Federal funds rates.

IV. Simulation Results and Conclusions

The model described in the previous three sections was used to simulate daily Federal Reserve open market purchases and sales and banking system response for fifty weeks beginning March 23, 1967. The structure of regression coefficients in Table 2 was employed to generate Federal Reserve open market decisions. The values of the independent variables $RDIF_t$, FR_t and $TRGR_t$ employed, however, were the values generated endogenously by the simulation model, after the start-up period. In terms of the desiderata stated at the outset of this paper, changes in earning assets and reserve positions were endogenous both to individual banks and to the system. Changes in demand deposits were exogenous to individual banks but endogenous to the system. Finally, the endogeneity of Federal Reserve open market operations was made explicit. The total system thus was closed insofar as possible.

Several observations regarding the "feedbacks" of the banking system of Federal Reserve actions may be highlighted.

1) As the banking system's demand for borrowed funds is made manifest in terms of rising rates on Federal funds, the Federal Reserve is led to reduce the rate at which it purchases government securities. Thus, expansive actions of the banking system tend to be self-limiting when considered within the total system. 2) As free reserves rise, reflecting declines in bank borrowing from the Federal Reserve or, alternatively, a rise in unborrowed reserves relative to required reserves, the Federal Reserve reduces its pace of purchases. If this were not the case, the total system would "blow up." 3) If the reserve base has been expanding in the recent past, it will tend to expand further in the current period. This reflects the inertial characteristics of the system.

The findings reported below use statistics for R^2 along with graphs to summarize the relationship between simulation-generated and actual time-series. We employ R^2 not in order to suggest any probabilistic interpretation but because it is an easily understood measure. On the other hand, squaring deviations may not be the most appropriate technique, given the capacity of Monte Carlo model specification to increase or decrease random movements. Therefore we have provided graphs of simulation and actual time-series, in terms of four-week averages, to suggest the degree to which the simulation captured—or failed to capture—major movements in endogenous variables.

We would conclude that these results, albeit based on primitive model designs, support the concept of a disaggregated approach to studying banking system—Federal Reserve interactions:

TABLE 3—SUMMARY OF REGRESSIONS OF SIMULATED VARIABLES ON HISTORIC VARIABLES

Variable	R^2
Open-market purchases and sales (OMP_t)	.21
Federal funds rate (r_{3t})	.51
Excessive Reserves (RX_t)	.02
Borrowings from Federal Reserve (RB_t)	.02
Total Reserves (TR_t)	.81
Demand Deposits (DD_t)	.84

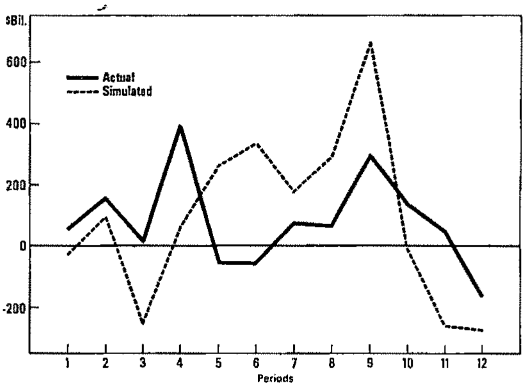


FIGURE 1. OPEN MARKET PURCHASES AND SALES. DATA ARE FOR FOUR WEEKS BEGINNING THE WEEK OF MARCH 23, 1967.

1. With the exception of excess reserves and borrowings, the simulations track history quite well. Given our specification the excess reserves result was to be expected; in fact we would have been hard-pressed to account for a good fit. The results with borrowings were disappointing, but we believe they could be substantially improved with the use of empirically estimated micro unit decision functions.

2. There is a pronounced tendency for simulated values of the Federal funds

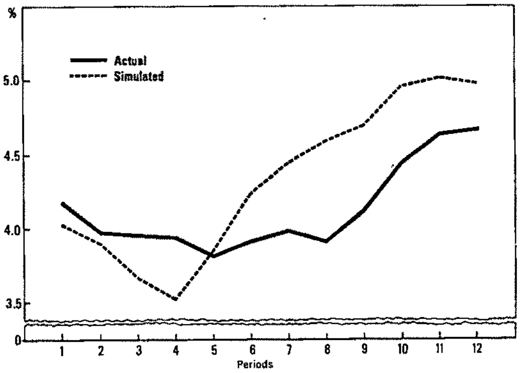


FIGURE 2. FEDERAL FUNDS RATE. DATA ARE FOUR-WEEK AVERAGES BEGINNING THE WEEK OF MARCH 23, 1967.

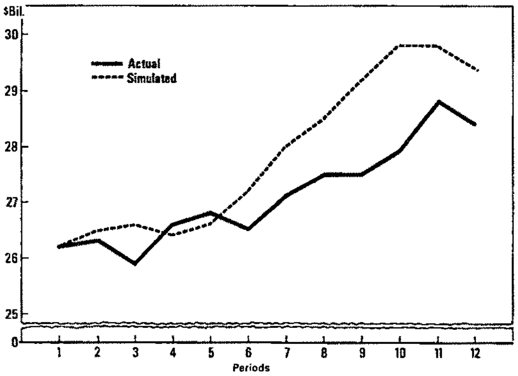


FIGURE 3. TOTAL RESERVES. DATA ARE FOUR-WEEK AVERAGES BEGINNING THE WEEK OF MARCH 23, 1967. NOTE: $2.8 BILLION HAS BEEN ADDED TO ACTUAL SERIES TO ADJUST FOR SCALE DIFFERENCES IN INITIAL CONDITIONS.

rate, open market purchases, borrowings, and, hence, increases in total reserves and demand deposits, to exceed historic values during the middle weeks of the simulation. We suspect this divergence is attributable to the assumption of unchanged bank decision rules over the fifty-week period. That is, had the rules specified less aggressive behavior during this period, the banks would have then put themselves under less of the reserve position pressure which forces up the Federal funds rate and member bank borrowing. There would also thus have been less such pressure for open market operations to offset via more liberal purchases of securities.[7] Mechanically, it would have been possible to experiment with the decision rules to improve the fit. We chose not to do this because the simulations as run call attention to the major need for empirical work on at least a representative sample of real banks. For this, it would be useful to have individual bank data on a daily basis over an extended period to time. Such data are not currently available, and substantial effort would be required to construct the appropriate data matrix.

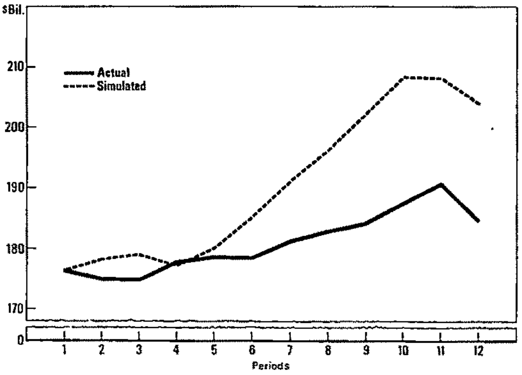


FIGURE 4. DEMAND DEPOSITS. DATA ARE FOUR-WEEK AVERAGES BEGINNING THE WEEK OF MARCH 23, 1967. NOTE: $46.2 BILLION HAS BEEN ADDED TO ACTUAL SERIES TO ADJUST FOR SCALE DIFFERENCES IN INITIAL CONDITIONS.

3. It is difficult to resist the comment that the stability of the model when closed suggests something about the real world. This suggests that Federal open market strategies involving less concern for short-run money market phenomena should also be simulated. Rather than attempt this in the current paper we would first call for more empirical content to the model. Then we could have at hand a tool available for monetary policy simulation experiments generally. Existing econometric models are quarterly; aside from the defects mentioned early in this paper, they are inherently unsatisfactory vehicles for monetary policy simulations because they cannot deal with the fact that policy is formulated and conducted continuously. A more fully developed model of the kind described in this paper could provide a useful alternative to the quarterly models, because it is designed in the same time-frame context within which monetary poicy, for better or worse, is conducted.

REFERENCES

L. C. Andersen and K. M. Carlson, "A Monetarist Model for Economic Stabilization,"

---

[7] The closed nature of the simulation model also serves to cumulate any real-world-model divergencies initially produced by unchanged bank decision rules.

Fed. Reserve Bank St. Louis Rev., Apr. 1970, *52*, 7–25.

——— and M. O. Rigg, "1967—A Year of Constraints on Monetary Management," *Fed. Reserve Bank St. Louis Rev.*, May 1968, *50*, 6–20.

K. Brunner and A. H. Meltzer, "Some Further Investigation of Demand and Supply Functions for Money," *J. Finance*, May 1964, *19*, 240–83.

W. R. Bryan and W. T. Carleton, "Short-Run Adjustments of an Individual Bank," *Econometrica*, Apr. 1967, *35*, 321–47.

F. de Leeuw and E. M. Gramlich, "The Channels of Monetary Policy," *Fed. Reserve Bull.*, June 1969, *60*, 472–91.

J. M. Guttentag, "The Strategy of Open Market Operations," *Quart. J. Econ.*, Feb. 1966, *80*, 1–30.

J. L. Jordan and C. E. Ruebling, "Federal Open Market Committee Decisions in 1968—A Year of Watchful Waiting," *Fed. Reserve Bank St. Louis Rev.*, May 1969, *51*, 6–15.

G. G. Kaufman and R. D. Laurent, "Simulating Policy Strategies on the FRB-MIT Model Under Two Alternative Monetary Policy Regimes," unpublished paper, Fed. Reserve Bank Chicago, Oct. 1969.

A. J. Meigs, *Free Reserves and the Money Supply*, Chicago 1962.

F. Modigliani, R. Rasche, and J. P. Cooper, "Central Bank Policy, the Money Supply and the Short-Term Rate of Interest," *J. Money, Credit, Banking*, May 1970, *2*, 166–217.

J. H. Wood, "A Model of Federal Reserve Behavior," in G. Horwich, ed., *Monetary Process Policy: A Symposium*, Homewood 1967, pp. 135–66.

Board of Governors, Federal Reserve System, *Federal Reserve Bulletin*, various issues.

Quarterly Economic Trends, Federal Reserve Bank of St. Louis, various issues.

The Federal Reserve and the Treasury: Answers to Questions from the Commission on Money and Credit, Englewood Cliffs 1963.

COMMUNICATIONS

Optimal Distributed Lag Responses and Expectations

By ROGER CRAINE*

Gradual adjustment in economic models can be divided into two categories; lags which the decision maker for an economic unit controls, and lags which he cannot control. Friction in the system, e.g., gestation periods, creates lags outside the decision maker's region of control. When the decision maker for an economic unit exerts no control over a lag, the best an econometrician can do is to analyze the economic process and attempt to measure the lag. With no specific information, unrestricted flexible measurement techniques benefit the econometrician by allowing the data to determine the form of the lag.

This paper examines lags which the decision maker fully controls. The speed of adjustment is subject to adjustment costs, but no friction. The decision maker can shorten, or eliminate, the lag if he is willing to pay the cost. Under these conditions, the optimizing decision simultaneously determines the profit maximizing levels of operation and the best adjustment path. If the econometrician can derive an analytic version of the optimal adjustment path he gains valuable information about the structure he intends to estimate.

Robert Eisner and Robert Strotz wrote the seminal paper on optimal distributed lag adjustment. They derived the optimal adjustment path and expressed it in the format of the familiar Koyck partial adjustment equation. Arthur Treadway and Robert Lucas built on their work making the objective function more general and expanding it to a vector of optimal adjustment paths. However, Lucas retained the stringent assumption made by Eisner and Strotz, that the independent variables maintain a constant path over the entire planning horizon. Treadway adopted the equivalent assumption of static expectations. Static expectations, or a constant path for the independent variables give adjustment equations with targets or "desired levels" that are *fixed* over time.

John P. Gould, Lance Taylor, and Peter Tinsley tackled models with changing expectations, or nonconstant paths of the independent variables. Taylor concentrated on a model with expectations driven by a particular Gauss-Markov process, while Gould solved a one-dimensional deterministic model for an optimal adjustment path. Tinsley solved a more general model which covers both Gould and Taylor's cases. He expressed the optimal adjustment equation in a partial adjustment format where the "target" depends on all the nonconstant future (or expected future) values of the independent variables.

This paper is primarily based on Tinsley's solution. The purpose is twofold: 1) to show the advantage in using an optimal adjustment specification and 2) to show the critical role that future values (or forecasts) play in determining the optimal adjustment path.

To demonstrate the advantage in an optimal adjustment specification I compare a derived optimal adjustment path with an *ad hoc* adjustment specification. The optimizing process places restrictions on the adjustment path and yields explicit functional forms which may be used to estimate the lag pattern more efficiently. In addition, the comparison makes explicit the implicit assumptions associated with *ad hoc* adjustment mechanisms and attributed to adjustment costs. For example, when *ad hoc* adjustment patterns violate the restrictions im-

---

* Federal Reserve System, Board of Governors. I wish to thank members of the Board's staff and especially Peter Tinsley, for their comments.

posed by the optimizing process, it implies the decision maker consciously follows a sub-optimal policy.

In the final section, I examine the impact of future values on the optimal adjustment equation and methods to approximate the future values. The target is extremely sensitive to different forecasting specifications. I discuss the strengths and weaknesses of three forecasting techniques. They range from naive static forecasts to a trajectory of forecasts for a nonstationary probability distribution.

Throughout the paper I use the firm's demand for quasi-fixed factors as an example of the adjustment process.

I. An *Ad Hoc* Adjustment Model

Econometricians segment the construction of *ad hoc* adjustment models into two independent operations. First they solve a static model for the equilibrium conditions. Then, the static model is "dynamized" by adding an *ad hoc* adjustment process which represents the retarding effects of the adjustment costs. The lag spreads the realization of the static equilibrium over time.[1]

To illustrate the process I derive the well-known Koyck partial adjustment equation using a static neoclassical model of a firm. A generalized objective function is used where the firm either maximizes profit or minimizes cost depending on whether the firm determines output, or whether it meets a given output constraint. The objective function is:

$$\underset{\max}{\pi} = \int_0^\infty v dt \tag{1}$$

$$v = \begin{cases} v_1(x, q \mid p, r, \delta) & \text{profit maximization} \\ v_2(x, \phi \mid q, r, \delta) & \text{cost minimization} \end{cases}$$

$$v = e^{-\delta t}\{pq - rx - \phi[q - f(x)]\}$$

where:

π = the present value of the profit stream
δ = the decision maker's discount rate
p = price of output
$r = (n\text{x}1)$ vector of factor rental rates
$x = (n\text{x}1)$ vector of factor inputs
ϕ = marginal cost
$f(x)$ = the production function
q = quantity

I make the conventional assumptions that (i) the firm operates in perfect competition in both the product and factor markets, (ii) the objective function is defined in the positive orthant, is strictly concave, and (iii) the production function has positive marginal products and passes through the origin.

The necessary conditions for maximization are:

$$(2) \qquad \begin{aligned} &\text{(a)}\ v_x = \phi f_x - r = 0 \\ &\text{(b)}\ v_\phi = q - f(x) = 0 \end{aligned}$$

The solution is a vector with $n+1$ elements. Factor inputs, x, constitute n elements. The remaining element is output q, or marginal cost ϕ. If the predetermined output constraint, which I call Q to distinguish it from output determined by the firm, equals the profit maximizing output q, then the two solutions coincide, i.e., if,

$$Q = q \quad \text{then,}$$

$$\phi = p \quad \text{and,}$$

$$x \text{ (cost minimizing)} = x \text{ (profit maximizing)}.$$

Notice the values that satisfy (2) are independent of time. The integral is maximized by maximizing at each point in time.

Solving equation (2a) for the static equilibrium level of factor inputs gives:[2]

$$x^s = \overset{-1}{f_x}(r, \phi)$$

The static equilibrium vector, x^s, serves as

---

[1] For an example of the two-step-construction process, see M. Ishaq Nadiri and Sherwin Rosen, pp. 459–61.

[2] The Jacobian of the production function never vanishes guaranteeing the unique inverse transformation f_x^{-1} exists. The Lagrangian multiplier depends on the model chosen,

$$\phi = \begin{cases} p & \text{profit maximization} \\ \phi(Q) & \text{cost minimization} \end{cases}$$

when the firm chooses its output level it equates marginal revenue, ϕ, to price, and the multiplier is easily eliminated. However, when the firm meets an output constraint, marginal costs, $\phi(Q)$, depend on the given output level.

the target in the partial adjustment equation. The *ad hoc* adjustment mechanism is:

$$(3)\quad x_{t} - x_{t-1} = \Delta x_{t} = L^{s}[x_{t}^{s} - x_{t-1}]$$

where:

$$L^{s} = \text{constant matrix}^{3}$$

or in distributed lag form,

$$(4)\quad x_{t} = L^{s} \sum_{i=0}^{t} [I - L^{s}]^{i} x_{t-i}^{s} \quad \text{for large t}$$

The simplicity and flexibility of the adjustment relation (3) make it very popular. The equation uses only observed values of the variables. Thus, the decision maker (or model builder) need not know or consider the future. In addition, a non-linear-analytic expression, reflecting the curvature of the production surface, serves as a target. The non-linear target might produce a difficult equation to estimate, but the procedure of separating the static-economic decision from the process determining the lag structure encourages the econometrician to pick an adjustment form convenient for estimation. With a little ingenuity one can usually justify a lag that leads to a transformed linear adjustment equation.[4]

Unfortunately, the same *ad hoc* specification that makes the adjustment model flexible and simple also makes it inefficient. The model is inefficient because it fails to incorporate all the available information.

To begin with, any internally consistent adjustment model (derived or *ad hoc*) must meet certain time independent restrictions,[5] e.g., accounting identities and physical constraints such as the production function. These restrictions hold at all times, including static disequilibrium positions. They reflect the overdetermination of a system subject to binding constraints. In a stochastic model the information in the constraint can be combined with the information in the sample to give a more efficient estimate. Failure to account for the interdependency not only wastes information, it results in the violation of a constraint, or throws all the compensating variation needed to meet the constraint in an omitted equation.

For example, when profits are maximized, output reflects any movement in a factor input not offset by movement in another factor,

$$(5)\quad \begin{aligned} \Delta q_{t} &= f_{x} \Delta x_{t} = f_{x} L^{s}[x_{t}^{s} - x_{t-1}] \\ &= \sum_{j}^{n} f_{x_j} \sum_{i}^{n} L_{jj}^{s} [x_{t_i}^{s} - x_{t-1_i}] \end{aligned}$$

When costs are minimized output is given and cannot absorb the compensating variation needed to meet the constraint. As a result, output must be invariant with respect to rental rate changes, i.e.,[6]

$$(\text{i})\quad q_{r_i} = F x_{r_i} L^{s} x_{t}^{s} + f_{x} L^{s} x_{r_i}^{s} = 0 \qquad i = 1, \ldots, n$$

where:

$$F = f_{xx}$$

Furthermore, since

$$(\text{ii})\quad \begin{aligned} q_{t} &= Q_{t}, \\ q_{Q} &= F x_{Q} L^{s} x^{s} + f_{x} L^{s} x_{Q}^{s} = 1 \end{aligned}$$

In practice non-linear restrictions are difficult to impose, however if the first-order conditions (2a) and (2b) are linear approximations or can be transformed to a linear form, then restrictions (i) and (ii) collapse to a linear vector which may be used to constrain a linear vector regression. Assume (2a) and (2b) can be written:[7]

---

[3] The lag structure could be more complicated without qualitatively changing the results that follow.

[4] For example, Nadiri and Rosen, p. 461, assume a Cobb-Douglas production function and a percentage adjustment equation yielding a log-linear relation for estimation. Koyck also uses this hypothesis on p. 69.

[5] William Brainard and James Tobin, and Nadiri and Rosen discuss the implications of dynamically consistent adjustment patterns in greater detail.

[6] I would like to thank Martin Ringo for pointing out an error in an earlier derivation of these restrictions. The derivation assumes a *known* constraint.

[7] A model with a Cobb-Douglas production function and the variables expressed in logs, e.g.,

$$x_{t}^{s} = B \begin{bmatrix} \log\ r \\ \log\ Q \end{bmatrix}_{t}$$

gives a set of log-linear first-order conditions (see

$$(2') \quad \begin{aligned} &(a)\ x_t^s = B\begin{bmatrix} r \\ Q \end{bmatrix}_t \\ &(b)\ q_t = \bar{f}_x x_t \end{aligned}$$

Linear first-order conditions imply,

$$F = 0$$

$$f_x = \bar{f}_x, \text{ a constant vector.}$$

Using the simplified constraints we can rewrite the partial adjustment equation as a linear vector equation subject to linear restrictions:[8]

$$(3') \quad x_t = L^s B\begin{bmatrix} r \\ Q \end{bmatrix}_t + [I - L^s]x_{t-1}$$

and,

$$\begin{aligned} \bar{f}_x x_t &= \bar{f}_x L^s B\begin{bmatrix} r \\ Q \end{bmatrix}_t + \bar{f}_x[I - L^s]x_{t-1} \\ &= Q_t \qquad\qquad\quad + \bar{f}_x[I - L^s]x_{t-1} \end{aligned}$$

Therefore, the restrictions are:

$$\bar{f}_x L^s B = [0, \ldots, 0, 1]$$

$$\bar{f}_x[I - L^s] = [0, \ldots, 0]$$

II. An Optimal Adjustment Equation

To derive more specification information about the adjustment path requires a more complete specification of the original model. Adjustment costs are an obvious addition.

$$(6)\ \pi = \int_0^\infty v dt$$

$$v = e^{-\delta t}\{pq - rx - a(x') - \phi[q - f(x)]\}$$

where:

$a(x')$ = adjustment cost function[9]

$x' = dx/dt$

I assume the adjustment cost function is strictly convex and passes through the origin.

The explicit recognition of adjustment costs makes the problem truly dynamic. The optimal solution simultaneously determines the adjustment process and the factor inputs. Each period's decision depends on every other period.

The modified objective function, equation (6), reaches an extreme when its Euler equations are satisfied. The Euler equations are:

$$(7) \quad \begin{aligned} &(a)\ \frac{dv}{dx} - \frac{d}{dt}\left(\frac{dv}{dx'}\right) = Ax'' - \delta a_{x'} + \phi f_x - r = 0 \\ &(b)\ \frac{dv}{d\phi} \qquad\qquad\quad = q - f(x) \qquad\qquad\quad = 0 \end{aligned}$$

where:

$$A = a_{x'x'}$$

The solution to equation (7) consists of $n+1$ *time paths*, n paths for the factor inputs, plus a path for output or marginal cost.

In general, analytic solutions to matrix differential equations only exist for linear equations with constant coefficients. Therefore, we assume a linear approximation adequately represents the Euler equations (7). The linear approximation trades an analytic solution to the dynamic model for a more precise functional specification. Actually the distortion is probably not large for short time intervals, since the linear approximation dominates in the neighborhood of the optimum path.[10]

Following Tinsley we write the optimal adjustment equation (which satisfies the linear approximation to the Euler equation) as:[11]

$$(8) \qquad x'(t) = L^*[x^*(t) - x(t)]$$

where:

Nadiri and Rosen). A model with a quadratic function and a linear constraint also yields linear first-order equations (see Henri Theil). If the first-order conditions are linear approximations then the constraints are only approximations.

[8] Nadiri and Rosen show these restrictions on p. 461. They use a log linear Cobb-Douglas production function. See McGuire et al. for an efficient method to utilize linear constraints in estimating a linear vector equation.

[9] In this section I use a continuous model, where the primed variable denotes a time derivative. The discrete analogue has similar properties, but leads to more complicated notation.

[10] See Raimond Struble, pp. 133–41, for the necessary conditions for linear dominance.

[11] A derivation of the solution appears in the Appendix.

$$
(9) \qquad
\begin{aligned}
x^*(\mathrm{t}) &= N^* \int_{\mathrm{t}}^{\infty} T_c e^{\lambda^+(\mathrm{t}-\xi)} T^c g(\xi) d\xi \\
N^* &= I - T_c[\lambda^-]^{-1}\lambda^+ T_c^{-1} \\
L^* &= -\, T_c \lambda^- T_c^{-1}
\end{aligned}
$$

The target, $x^*(\mathrm{t})$, and the adjustment matrix, L^*, are expressed in terms of the characteristic root matrices, $\{\lambda^+, \lambda^-\}$, the transformation matrices, $\{T_c, T^c\}$, and the forcing function, $g(\mathrm{t})$, of the linearized version of Euler equation (7).

$$
\begin{aligned}
&\overset{-}{\lambda}_{ii} < 0 && i = 1, \ldots, n\text{[12]} \\
&\overset{+}{\lambda}_{ii} = > \delta > 0 && i = n + 1, \ldots, 2n \\
&g(\mathrm{t}) = \overline{A}^{-1}[r(\mathrm{t}) - \bar{f}_x \phi(\mathrm{t}) + \bar{\phi}[\bar{f}_x + \overline{F}\bar{x}] - \bar{r}] \\
&\qquad = 0 \\
&C = -\, \bar{\phi}\overline{A}^{-1}\overline{F} = -\, T_c[\lambda^-\lambda^+]T_c^{-1} \\
&T^c = [\lambda^- - \lambda^+]^{-1} T_c^{-1}
\end{aligned}
$$

The bar notation indicates a constant vector or matrix (the variables evaluated at the expansion point). Notice the target in equation (8) is actually a trajectory given by the integral equation (9).

Integrating equation (8) gives the solution in distributed lag form:[13]

$$
(10) \qquad
\begin{aligned}
x(\mathrm{t}) = &\int_0^{\mathrm{t}} e^{-L^*(\mathrm{t}-\tau)} L^* x^*(\tau) d\tau \\
&+ e^{-L^*\mathrm{t}} x(0) \\
&\cong L^* \int_0^{\mathrm{t}} e^{-L^*(\mathrm{t}-\tau)} x^*(\tau) d\tau \quad \text{for large t}
\end{aligned}
$$

Comparing the optimal-adjustment equations (8) and (10) with a consistent version of their *ad hoc* counterparts, equations (3) and (4), reveals several differences.

The fundamental difference lies in the formulation of the two targets. The integral target in the optimal-adjustment equation, $x^*(\mathrm{t})$, is a distributed lead discounting all the future values of the independent variables, $g(\mathrm{t})$. The integral recognizes future events and smooths jagged changes in $g(\mathrm{t})$. In contrast, the static target, $x^s(\mathrm{t})$, depends only on current values. The static target shifts abruptly mirroring any sharp changes in the independent variables since it ignores the future.

A less significant difference between the two models arises because the optimal adjustment equation (9) is a linear approximation of the true path. The linear approximation is a good approximation in the neighborhood of the true solution. The *ad hoc* adjustment equation has no fixed functional form by definition. However, the model builder must choose non-linear adjustment forms carefully if he characterizes the adjustment process as optimal. Non-linear expressions carry implications for other portions of the model which he may not intend. For example, Nadiri and Rosen (p. 461) assume a log-linear adjustment relation which implies a multiplicative objective function.

In other ways the equations are very similar. The adjustment equations (3) and (8) have constant adjustment matrices, L^s and L^*.[14] Written in distributed-lag form these adjustment matrices are used to discount past values of the targets. The speed of adjustment in the optimal-adjustment model varies directly with the discount rate and the rate of change in the slope of the production surface, but inversely with respect to the rate of change in the adjustment cost surface.[15] Of course, the speed of adjustment in the *ad hoc* model also depends on these arguments, but not in a well-defined analytic form.

The major distinction between the models boils down to the treatment of future values in the target. In the special case when the time path of the future values is a constant the target, $x^*(\mathrm{t})$, collapses to a constant. The constant is the linearized static target.

---

[12] Nissan Liviatan and Paul Samuelson discuss the saddlepoint property of the roots. The roots in our system correspond to the matrix version of the roots they give in equation (9) when $F_{12}=0$.

[13] Tinsley calls the solution written in distributed lag form a "double Koyck." He also gives a transformation to express the solution in "double partial adjustment" form. In partial-adjustment form the equation can be easily estimated, see Tinsley (1970b, p. 407).

[14] The constant adjustment matrix follows directly from the linearization of the Euler equations.

[15] See the Appendix, equation (A3). Notice that L^* is diagonal only if C is diagonal.

Integrating the target over the constant function g and substituting terms gives:[16]

$$\bar{x} = [\bar{\phi}\overline{F}]^{-1}\overline{A}\bar{g} \tag{11}$$

The partial-adjustment equation for a fixed target model is:

$$x'(\mathrm{t}) = L^*[\bar{x} - x(\mathrm{t})] \tag{12}$$

Notice the target x does not change over time. Written in distributed-lag form the weakness of fixed target models for empirical work becomes more transparent.

$$\begin{aligned} x(\mathrm{t}) &= e^{-L^*\mathrm{t}}[x(0) - \bar{x}] + \bar{x} \\ &\cong \bar{x} \qquad \text{for large t} \end{aligned} \tag{13}$$

that is, after sufficient time has elapsed equation (13) collapses to a constant.[17]

The justification for estimating fixed-target models rests on risk. If the decision maker has static expectations in a risky world, he must continually revise his target.

III. Optimal Adjustment with Risk

The simplest way to add risk to the model is to assume the vector path of independent variables is a random vector. I assume the vector is normally distributed, but not stationary.[18]

$$g(\mathrm{t}) = \epsilon(\mathrm{t})$$

where:

$$\epsilon(\mathrm{t}) \sim N(\mu(\mathrm{t}), \Sigma(\mathrm{t}))$$

Given risk, the decision maker tries to maximize the expected value of the objective function. As a result, the form of the decision maker's expectations function plays a critical role in determining the nature of the solution. Below I present three forms for an expectations function. The dramatic changes they produce in the target illustrate the importance of choosing a realistic expectations function in a model that efficiently uses the information available.

The most naive forecaster is a static forecaster. I call it a symmetric myopic forecast.

$$E[g(\tau \mid \mathrm{t})] = g(\mathrm{t}); \qquad \tau > \mathrm{t}$$

He assumes the current values of the independent variables remain constant forever. The symmetric myopic's expectation function adds a time dimension to the target in equation (12).

$$x' = L^*[\bar{x}(\mathrm{t}) - x(\mathrm{t})] \tag{12$'$}$$

where:

$$E[x(\mathrm{t})] = \bar{x}(\mathrm{t})$$

Mathematically equation (12′) is indistinguishable from a linearized version of the consistent *ad hoc* adjustment equation (3). However, the formulation of equation (12′) exposes the static expectations assumption. The *ad hoc* adjustment equation (3) appears to use only observed values. Whereas, in equation (12′) although only observed values explicitly appear the static expectations implicitly incorporate all the future values by assuming that they remain *constant*. Of course, even if the naive-static forecaster correctly predicts a stationary distribution the target still varies randomly about a fixed mean. As a result, the static forecaster pays adjustment costs to compensate for random variations.

A slightly more sophisticated forecaster might be called a nonsymmetric myopic. The nonsymmetric myopic remembers he has made errors in the past, but he still believes the underlying distribution is stationary. His problem is to find a consistent estimate of the mean, μ. Assume he makes a new decision at discrete intervals. By applying a Bayesian recurrence relation, the forecaster can improve his current estimate by remembering his past predictions. His Bayesian estimate of μ is given by:

---

[16] The derivation of the stationary target is easier to see in equation (7). Set, $x' = x'' = 0$, solve and linearize giving, $x = [\bar{\phi}\overline{F}]^{-1}\overline{A}g$.

[17] Most of the models in growth theory have fixed targets. See Frank Hahn or Edmund Phelps.

[18] When the first-order conditions are linear it is possible to work out the optimal decision process independently of the expectation formation process, e.g., see Theil. As a consequence, conditional expectations can be substituted for the future values of the deterministic variables in equations (8), (10), and (12). The certainty equivalence property does not require a normal distribution.

$$P(\mu_t \mid \epsilon_t, \mu_{t-1}) \propto J(\epsilon_t \mid \mu, \Sigma) P(\mu_{t-1})$$

where:

$P(\mu_t)$ = the posterior distribution of μ at time t
$J(\epsilon_t|\mu,\Sigma)$ = likelihood function at time t
$P(\mu_{t-1})$ = prior distribution for μ, which is the previous period's posterior distribution

$$E[\mu_t] = (t\mu + \mu_0)/t + 1$$
$$\operatorname{var}[\mu_t] = \Sigma/t + 1$$

The Bayesian-updating procedure builds in all of the past information. The nonsymmetric myopic's expectation function is:[19]

$$E[g(\tau \mid t)] = \frac{\epsilon(t) + t\mu(t-1)}{(t+1)}; \quad \tau \geq t$$

The Bayesian procedure provides a consistent estimate for a stationary distribution.[20] In contrast, the variance of the symmetric myopic's forecast remains constant.

Finally, consider a decision maker who forecasts the mean of a nonstationary distribution. His expectations function can be represented as:

$$E[g(\tau \mid t)] = u(\tau); \qquad \tau \geq t$$

The nonstationary target takes the form of the distributed lead given in equation (9) with conditional expectations substituted for the deterministic forward values.

The forward stream of expectations raises the problem of how to summarize the target in a form suitable for estimation. Rewriting the partial-adjustment equation (10) in discrete notation gives:

$$x_t - x_{t-\Delta} = \Delta L^{**}[E(x_t^{**}) - x_{t-\Delta}] \tag{13}$$

where:

$$\Delta = \frac{t-0}{n}$$

$$E(x_t^{**}) = N^{**} \sum_{i=t}^{\infty} D^{(n-i)\Delta} E(g_{i\Delta})$$

The matrices are defined so that,

$$x'(t) = \lim \Delta \to 0 \left[\frac{x_t - x_{t-\Delta}}{\Delta}\right]$$

Tinsley proposes differencing equation (13 again to put it in a double partial adjustmen form.[21] Setting,

$$\Delta = 1$$

and subtracting,

$$KE(x_{t+1}) = L^{**}N^{**} \sum_{i=t+1}^{\infty} D^{(t+2-i)} E(g_i) + K[I - L^{**}]x_t$$

where:

$$K = L^{**}N^{**}D[L^{**}N^{**}]^{-1}$$

gives,

$$x_t = H^{-1}\{L^{**}N^{**}g_t + KE(x_{t+1}) + [I - L^{**}]x_{t-1}\} \tag{14}$$

where:

$$H = [I + K - KL^{**}]$$

The differencing operation transforms th target from a distributed lag form to partial adjustment form. The observed current vector of independent variables $g(t)$, and the expected value of the dependent-variables vector in the next period $E(x_{t+1})$, replace the discounted infinit stream of variables which make up the targe $E(x_t^{**})$. Unfortunately, finding a vector t represent $E(x_{t+1})$ is not a trivial task Properly interpreted, equation (14) repre sents a dynamic programming statement of the original optimization problem. The vector $E(x_{t+1}^{**})$ lies on the *expected* optimal path. It is the optimal vector for period t+1 given the conditional expectations of the entire time path for the independent variables at time t. The vector $E(x_{t+1})$ reflects the expected path of the independent variables in the same manner that the target $E(x_t^{**})$ did. Substituting actual values for $E(x_{t+1})$ assumes perfect foresight. The variable $E(x_{t+1})$ must contain any forecast errors, or the coefficient estimates will be biased. Anticipations data do contain forecast errors and might be a good candidate for $E(x_{t+1})$.

---

[19] See Dennis Lindley, p. 3, or Tinsley (1970a, p. 23).

[20] For a similar updating method based on a Kalman filter of an autocorrelated system, see Taylor. I assume the initial likelihood and prior distribution have the same variance. Assuming different variances changes the weights on past observations.

[21] Differencing equation (13) is equivalent to dif ferentiating the continuous partial adjustment equatio (10). Of course, differentiating equation (10) gives th Euler equation. Equation (14) is the rearranged dis crete analogue of the Euler equation. The discret notation helps to emphasize the influence of futur events.

In any event, optimal-adjustment models depend in a fundamental way on future values. The dependence should be explicitly recognized and not cavalierly dismissed with the assumption of static expectations. Nonstationary functions and proxy variables for $E(x_{t+1})$ need to be tested.

APPENDIX

The solution technique presented here sketches the method followed by Tinsley (1970a, pp. 14–15). Linearizing the Euler equations (7) gives a two-point boundary value problem consisting of n second-order linear differential equations.

$$\begin{aligned} &\bar{A}x''(\mathrm{t}) - \delta\bar{A}x'(\mathrm{t}) + \bar{\phi}\bar{F}x(\mathrm{t}) \\ \text{(A1)}\quad &- [r(\mathrm{t}) - \bar{f}_x\phi(\mathrm{t}) + \bar{\phi}[\bar{f}_x + \bar{F}\bar{x}] - \bar{r}] \\ &= 0 \end{aligned}$$

I assume the constant expansion vector lies on the optimal path. Notice,

$$\bar{x} = \text{constant vector} \Rightarrow \bar{x}' = \bar{x}'' = 0$$

Substituting, $y(\mathrm{t}) = x'(\mathrm{t})$ we can rewrite equation (A1) as a set of $2n$ first-order differential equations.

$$\text{(A2)}\quad \begin{bmatrix} y'(\mathrm{t}) \\ x'(\mathrm{t}) \end{bmatrix} = \begin{bmatrix} \delta I & C \\ I & 0 \end{bmatrix} \begin{bmatrix} y(\mathrm{t}) \\ x(\mathrm{t}) \end{bmatrix} + \begin{bmatrix} g(\mathrm{t}) \\ 0 \end{bmatrix}$$

Recalling the definitions,[22]

$$T^c = [\lambda^- - \lambda^+]^{-1} T_c^{-1}$$

$$-T_c^{-1} C T_c = -\lambda_c = \lambda^+\lambda^-$$

It is easy to verify that the following transformation diagonalizes the characteristic matrix:

$$\text{(A3)}\quad \begin{bmatrix} T^c & -\lambda^+ T^c \\ -T^c & \lambda^- T^c \end{bmatrix} \begin{bmatrix} \delta I & C \\ I & 0 \end{bmatrix} \begin{bmatrix} T_c\lambda^- & T_c\lambda^+ \\ T_c & T_c \end{bmatrix}$$

$$= \begin{bmatrix} \delta\lambda - 2\lambda^+\lambda^- & -[\lambda^+]^2 + \delta\lambda^+ + \lambda_c \\ [\lambda^-]^2 - \delta\lambda^- - \lambda_c & -\delta\lambda^+ + 2\lambda^+\lambda^- \end{bmatrix}$$

$$[\lambda^- - \lambda^+]^{-1} = \begin{bmatrix} \lambda^- & 0 \\ 0 & \lambda^+ \end{bmatrix}$$

The general solution for $x(\mathrm{t})$ is:

$$\begin{aligned} x(\mathrm{t}) &= T_c e^{\lambda^- \mathrm{t}} k_1 + T_c e^{\lambda^+ \mathrm{t}} k_2 \\ \text{(A4)}\quad &+ \int_0^{\mathrm{t}} [T_c e^{\lambda^-(\mathrm{t}-\tau)} T^c \\ &- T_c e^{\lambda^+(\mathrm{t}-\tau)} T^c] g(\tau) d\tau \end{aligned}$$

To get the particular solution we must find the vector of constants $[k_1, k_2]$ that satisfies the $2n$ boundary conditions. The boundary conditions are:

$$\text{(Ai)}\quad x(0) = T_c k_1 + T_c k_2$$

$$\begin{aligned} &\lim_{\mathrm{t}\to\infty} e^{-\delta \mathrm{t}} \left\{ \frac{dv}{dx'(\mathrm{t})} = \bar{A}x'(t) \right. \\ &= \bar{A} \Big[T_c\lambda^- e^{\lambda^- \mathrm{t}} k_1 + T_c\lambda^+ e^{\lambda^+ \mathrm{t}} k_2 \\ \text{(Aii)}\quad &+ \int_0^{\mathrm{t}} (T_c\lambda^- e^{\lambda^-(\mathrm{t}-\tau)} T^c \\ &\left. - T_c\lambda^+ e^{\lambda^+(\mathrm{t}-\tau)} T^c) g(\tau) \mathrm{d}\tau \Big] \right\} = 0 \end{aligned}$$

The boundary condition (Ai) specifies an $(n\mathrm{x}1)$ vector of initial factor inputs. Condition (A2) is the transversality condition which gives an $(n\mathrm{x}1)$ vector of terminal conditions.

Solving (Ai) and (Aii) simultaneously and substituting the results for $[k_1, k_2]$ in the general solution gives:

---

[22] When the roots are *not* distinct, define:

$$\lambda^+ T^c = [\lambda^- - \lambda^+]^{-1} \lambda^+ T_c^{-1}$$

$$\lambda^- T^c = [\lambda^- - \lambda^+]^{-1} \lambda^- T_c^{-1}$$

$$\text{(A5)}\quad x(\mathrm{t}) = T_c e^{\lambda^- \mathrm{t}} T_c^{-1}[x(0) - \bar{x}(0)] + \bar{x}(\mathrm{t})$$

where the steady state trajectory,

$$\bar{x}(\mathrm{t}) = \int_0^{\mathrm{t}} T_c e^{\lambda^-(\mathrm{t}-\tau)} T^c g(\tau)d\tau + \int_{\mathrm{t}}^{\infty} T_c e^{\lambda^+(\mathrm{t}-\tau)} T^c g(\tau)d\tau$$

Differentiating equation (A5) with respect to time gives the partial adjustment equation (8). The solution (A5) is optimal and unique.

REFERENCES

W. C. Brainard and J. Tobin, "Pitfalls in Financial Model Building," *Amer. Econ. Rev. Proc.*, May 1968, *58*, 99–122.

R. M. Coen and B. G. Hickman, "Constrained Joint Estimation of Factor Demand and Production Functions," *Rev. Econ. Statist.*, Aug. 1970, *52*, 287–300.

R. Eisner and R. H. Strotz, "Determinants of Business Investment," in D. B. Suits et al., eds., *Impacts of Monetary Policy, Commission on Money and Credit*, Englewood Cliffs 1963, pp. 369–82.

J. P. Gould, "Adjustment Costs in the Theory of Investment of the Firm," *Rev. Econ. Stud.*, Jan. 1968, *35*, 47–55.

F. H. Hahn, "Some Adjustment Problems," *Econometrica*, Jan. 1970, *38*, 1–18.

L. M. Koyck, *Distributed Lags and Investment Analysis*, Amsterdam 1954.

D. V. Lindley, *Introduction to Probability Theory from a Bayesian Viewpoint, Part Two, Inference*, Cambridge 1965.

N. Liviatan and P. A. Samuelson, "Notes on Turnpikes: Stable and Unstable," *J. Econ. Theor.*, Dec. 1969, *1*, 454–75.

R. E. Lucas, Jr., "Optimal Investment Policy and the Flexible Accelerator," *Int. Econ. Rev.*, Feb. 1967, *8*, 78–85.

T. W. McGuire et al., "Estimation and Inference for Linear Models in which Subsets of the Dependent Variable are Constrained," *J. Amer. Statist. Ass.*, Dec. 1968, *63*, 1201–13.

M. I. Nadiri and S. Rosen, "Interrelated Factor Demand Functions," *Amer. Econ. Rev.*, Sept. 1969, *59*, 457–71.

M. Nerlove, "On Lags in Economic Behavior," Center for Mathematical Studies in Business and Economics, University of Chicago, Report No. 7109, Feb. 1971.

E. S. Phelps, *Golden Rules of Economic Growth*, New York 1966.

R. A. Struble, *Nonlinear Differential Equations*, New York 1962.

L. Taylor, "The Existence of Optimal Distributed Lags," *Rev. Econ. Stud.*, Jan. 1970, *37*, 95–106.

H. Theil, *Optimal Decision Rules for Government and Industry*, Amsterdam 1964.

P. A. Tinsley, (1970a) "On Optimal Dynamic Adjustment of Quasi-Fixed Factors," Special Studies Paper #9, Federal Reserve Board, 1970.

———, (1970b) "On Ramps, Turnpikes, and Distributed Lag Approximations of Optimal Intertemporal Adjustment," *Western Econ. J.*, Dec. 1970, *8*, 397–411.

A. B. Treadway, "Adjustment Costs and Variable Inputs in the Theory of the Competitive Firm," *J. Econ. Theor.*, Dec. 1970, *2*, 329–48.

The Theory of Tariffs, Taxes, and Subsidies: Some Aspects of the Second Best

By TRENT J. BERTRAND AND JAROSLAV VANEK[*]

The purpose of this paper is twofold. First, we establish three theorems defining the relationship between the gross and net production possibilities sets when production possibilities are bounded by the fixed supply of primary factors in the economy. We prove that the normal assumption of a strictly convex efficient surface on the gross production possibility set yields a strictly convex efficient surface on the net production possibility set. Furthermore, it is shown that perfect competition will lead to production on the efficient surface of the gross production possibility set where the marginal rate of transformation is equated to the ratios of unit values added and that the corresponding solution in terms of net outputs is characterized by equality of the marginal rate of transformation on the efficient surface of the net production possibility surface and the ratio of product prices. These results of profit maximizing behavior in a perfectly competitive economy provide constraints to the normative problems set out in Sections II and III.

The second and principal purpose of this paper is to examine the second best solutions corresponding to prescribed levels of some tariffs, taxes, or subsidies and variable levels of others. The interest of such an examination is perhaps obvious. Governments often are under obligation—owing to domestic pressure groups, international agreements, or legal commitments—to preserve tariffs or subsidies on some commodities. If this is so, our analysis then tells us what levels of protection should be selected for those commodities where no rigid commitments are present.

Based on a model with intermediate products, our analysis is also relevant to the newly developing general equilibrium theory of effective protection. By and large, concern in recent papers has focused on questions of the quantitative impact of effective protection.[1] Questions of second best for which a general equilibrium approach is so well suited have largely been neglected.

In Section II, it is proved that under certain conditions a welfare maximizing policy requires the elimination of extreme distortions which are within the control of policy makers when the existence of distortions outside their control prevent the attainment of a first best solution. In Section III, it is shown that the consumption possibilities facing a community may be increased by a similar elimination of extreme distortions in a model where the distortions are defined in terms of *effective* or value added distortions rather than nominal distortions.

I. The Production Possibility Sets

The economy consists of n commodities, X_i $(i=1 \ldots n)$, and m factors, v_h $(h=1 \ldots m)$. Technology is defined by production functions permitting substitution of primary factors in a neoclassical manner and the input of intermediate products in fixed proportions, that is;

$$(1) \quad X_i = \min\left[f^i(v_1^i, \ldots, v_m^i), \frac{X_1^i}{a_{1i}}, \ldots, \frac{X_n^i}{a_{ni}}\right]$$

where a_{ji} $(j=1, \ldots, n)$ is the fixed amount of commodity j used in the production of one unit of commodity i ($a_{ij}=0$ if $i=j$), v_h^i is the amount of the hth factor used in production of the ith commodity, and X_j^i is the total amount of the jth commodity used in producing the ith commodity.

---

[*] Assistant professor at the Johns Hopkins University and professor at Cornell University, respectively. We are indebted to Bela Balassa, Frank Flatters, Peter Newman, Hugh Rose, Alan Kirman, and a referee for helpful comments. Tatsuo Hatta has been especially helpful in several discussions of the methodology and results of this study, while Hajime Hori is responsible for pointing out several errors in an earlier draft.

[1] See, for instance, W. M. Corden, V. K. Ramaswami and T. N. Srinavasan, Roy Ruffin, Balassa and S. F. Guisinger. For an excellent exception dealing with the second best with different constraints than studied here, see A. H. H. Tan.

We assume that the effective constraint on production is always the availability of primary factors,[2] so that (1) may be rewritten as (1′):

$$X_i = f^i(v_1^i, \ldots, v_m^i) \tag{1′}$$

The gross production possibility set T,

$$T(X_1, \ldots, X_n) \leq 0 \tag{2}$$

is bounded due to a fixed endowment of primary factors. The relationship between gross and net outputs is defined in relation (3);

$$[I - A]X = x \tag{3}$$

where I is the identity matrix, A is the nxn matrix of input-output coefficients, X is a vector of gross outputs, and x is a vector of net outputs. The net production possibility set τ is therefore[3]

$$\tau\left(X_1 - \sum_j a_{1j}X_j, \ldots, X_n - \sum_j a_{nj}X_j\right) = \tau(x_1, \ldots, x_n) \leq 0 \tag{4}$$

and is obtained by subtracting outputs used as intermediate inputs from the gross outputs.

Preliminary to our analysis of protection, we establish three theorems concerning the gross and net production possibility sets.

THEOREM 1. *If the gross production possibility set is strictly convex, then the net production possibility set will be strictly convex.*

PROOF:

Choose any two points X' and X'' in the gross output set T and let X be a strictly convex linear combination of these points. Then by the strict convexity of T, we can choose $\overline{X}$ such that;

$$\overline{X} > X = \rho X' + (1 - \rho)X'' \tag{5}$$

where $0<\rho<1$, and $X \in T$. Premultiplying the right-hand side of inequality (5) by $[I-A]$, and by definition of τ, we have:

$$x = \rho x' + (1 - \rho)x'' \tag{6}$$

where $x \in \tau$, and x, x', and x'' are the net output vectors corresponding to X, X', and X'', respectively. Let δ be a strictly positive vector of arbitrarily small net outputs. Then τ is strictly convex if $(x+\delta) \in \tau$. But $(x+\delta) \in \tau$ if $(X+[I-A]^{-1}\delta) \in T$. By the Hawkins-Simon condition,[4] $[I-A]^{-1}$ is a matrix of positive elements so that we have

$$X + [I - A]^{-1}\delta > X \tag{7}$$

But since $[I-A]^{-1}\delta$ may be made arbitrarily small by the appropriate selection of δ, we choose δ such that,

$$\overline{X} > X + [I - A]^{-1}\delta \tag{8}$$

which means by (5), that $X+[I-A]^{-1}\delta \in T$ and therefore, we have $(x+\delta) \in \tau$.

THEOREM 2:[5] *Profit maximizing behavior with perfectly competitive product and factor markets will lead to production on the gross production possibility surface where the marginal rate of transformation is equated to the ratio of values added.*

PROOF:

Situated on the boundary of T, we consider the possibility of reallocating one unit of any factor of production from the ith to the jth industry. The total cost of production for the jth industry is;

$$TC_j = \sum_h w_h v_h^j + \sum_i P_i^d a_{ij} X_j \tag{9}$$

and for the ith industry;

$$TC_i = \sum_h W_h v_h^i + \sum_j P_j^d a_{ji} X_i \tag{10}$$

where P_i^d is the domestic price of the ith commodity. Differentiating (9) and (10) with one factor v_* variable and noting that profit maximizing behavior under perfect competition implies marginal cost pricing, we obtain;

---

[2] This assumption has been used previously by Vanek (1963).

[3] Summations, unless otherwise specified, are over all commodities or factors.

[4] The Hawkins-Simon conditions assure that the economy is not self-consuming.

[5] A direct proof of the equality of relative prices to the marginal rates of transformation along τ is provided for the two-product case in Vanek (1963).

$$(11)\quad MC_j = \frac{dTC_j}{dX_j} = w_* \frac{dv_*^j}{dX_j} + \sum_i P_i^d a_{ij} = P_j^d$$

and

$$(12)\quad MC_i = \frac{dTC_i}{dX_i} = w_* \frac{dv_*^i}{dX_i} + \sum_j P_j^d a_{ji} = P_i^d$$

Since v_* is reallocated from the ith to the jth industry, $dv_*^j = -dv_*^i$ and since w_* is the same for both industries, we obtain from (11) and (12);

$$(13)\quad -\frac{dX_i}{dX_j} = \frac{T_j^*}{T_i^*} = \frac{P_j^d - \sum_i a_{ij}}{P_i^d - \sum_j P_j^d a_{ij}} = \frac{\pi_j^d}{\pi_i^d}$$

where $\pi_i^d = P_i^d - \sum_j P_j^d a_{ji}$ is the domestic value added (or net price) of commodity i. The slope defined in relation (13) is the marginal rate of transformation along a production possibility surface, T^*, constrained by the immobility of factors other than v_*. Since we are evaluating this slope at a point on the boundary of T which (by definition) cannot be intersected by T^*, it must be equal to the slope of the unconstrained production possibility surface, that is

$$(14)\quad \frac{T_j}{T_i} = \frac{T_j^*}{T_i^*} = \frac{\pi_j^d}{\pi_i^d}$$

THEOREM 3: *If the marginal rates of transformation of gross outputs are equated to the ratios of value added, the marginal rates of transformation between net outputs will be equated to the ratio of their prices.*

PROOF:

Differentiating equation (4) and setting all changes in gross output except for the ith and jth industry equal to zero, we obtain;

$$(15)\quad (\tau_i - \sum_j a_{ji}\tau_j)\, dX_i + (\tau_j - \sum_j a_{ij}\tau_i)\, dX_j = 0$$

$$(16)\quad -\frac{dX_i}{dX_j} = \frac{\tau_j - \sum_i a_{ij}\tau_i}{\tau_i - \sum_j a_{ji}\tau_j} = \frac{1 - \sum_i a_{ij}\frac{\tau_i}{\tau_j}}{\frac{\tau_i}{\tau_j} - \sum_j a_{ji}}$$

Equating (16) to π_j^d/π_i^d and solving for τ_i/τ_j, we obtain;

$$(17)\quad \frac{\tau_i}{\tau_j} = \frac{P_i^d}{P_j^d}$$

The correspondence between the net output and gross output production possibility sets defined by these three theorems not only proves useful in this paper but makes possible the extension of many theorems concerning the effects of nominal tariffs (which distort prices) on net outputs to the effect of effective tariffs (which distort values added) on gross outputs.

II. Distortions through Tariffs and Subsidies on Trade[6]

We now consider the conditions for improving community welfare in a small country facing fixed terms of trade when the first best solution of zero or equal tariffs and subsidies is not possible either in the sense that all tariffs cannot be immediatly equated due to time constraints involved in changing commercial policy or because some tariffs cannot be altered because of political infeasibility. It is assumed in the analysis that tariff revenue is redistributed to the community.

With the nth commodity having an extreme distortion $(t_n >$ or $< t_i)$, policy authorities seek to improve community welfare by varying t_n while holding all other tariffs constant. It is assumed that the preferences of the community are represented by a convex nonintersecting preference map given by relation (18);

---

[6] A simplification of an earlier treatment of this section was suggested by Tatsuo Hatta.

$$u = u(C_1, \ldots, C_n) \tag{18}$$

where C_i is the consumptions of the ith commodity.

The budget constraint facing the community with tariff revenue redistributed is given in relation (19);

$$\begin{aligned} \sum_i P_i C_i &= \sum_i P_i X_i - \sum_i \sum_j P_j a_{ij} X_j \\ &= \sum_j P_i X_i \end{aligned} \tag{19}$$

where P_i is the fixed international price of the ith commodity, and which shows that the value of the consumption bundle in equilibrium is constrained by the value of net output, both evaluated at world prices.

Net production possibilities are given in relation (20) and the conditions for profit maximization and consumer utility maximization are given in relations (20), (21), and (22), respectively;

$$\tau(x_1, \ldots, x_n) = 0 \tag{20}$$

$$\frac{\tau_i}{\tau_j} = \frac{P_i(1+t_i)}{P_j(1+t_j)} \quad (i=1, \ldots, n, \ j=1, \ldots, n) \tag{21}$$

$$\frac{u_i}{p_i}(1+t_i) = A \quad (i=1, \ldots, n) \tag{22}$$

where t_i is the percentage price distortion due to the imposition of a tariff or trade subsidy and A is the marginal utility of income.

By differentiating (18), we have the change in social welare as;[7]

$$du = \sum_i u_i dC_i \tag{23}$$

which, by relation (22) may be rewritten as;

$$\begin{aligned} du &= A \sum_i P_i(1 + t_i) dC_i \\ &= \Big\{ \sum_i P_i dC_i + \sum_i P_i t_i dC_i \Big\} \end{aligned} \tag{24}$$

which, by differentiating (19) and substituting in (24) gives;

$$\begin{aligned} du &= A \Big\{ \sum_i P_i dx_i + \sum_i P_i t_i dC_i \Big\} \\ &= A \Big\{ \sum_i P_i(1 + t_i) dx_i \\ &\qquad - \sum_i P_i t_i dx_i + \sum_i P_i t_i dC_i \Big\} \end{aligned} \tag{25}$$

Differentiating relation (20) and using relation (21), it is seen that

$$\sum_i P_i(1 + t_i) dx_i = 0 \tag{26}$$

which can be substituted into (25) to obtain

$$\begin{aligned} du &= A \Big\{ - \sum_i P_i t_i dx_i + \sum_i P_i t_i dC_i \Big\} \\ &= A \Big\{ \sum_i P_i t_i dE_i \Big\} \end{aligned} \tag{27}$$

where $E_i (= C_i - x_i)$ is the excess demand of the community for the ith commodity.

Letting e denote the commodity with extreme distortions, we have;

$$du = A \Big\{ t_e P_e dE_e + \sum_j P_j t_j dE_j \Big\} \tag{28}$$

where the j subscript denotes the $n-1$ other commodities. But by differentiating the budget constraint, we have;

$$P_e dE_e = - \sum_j P_j dE_j \tag{29}$$

which when substituted into (28) gives relation (30).

$$\frac{du}{dt_e} = A \left\{ \sum_j P_j (t_j - t_e) \frac{\partial E_j}{\partial t_e} \right\} \tag{30}$$

It defines the change in community welfare when t_e is varied but all other distortions are held constant. Equation (30) is central to our analysis.

First, we define commodity i as a complement of commodity j if excess demand for commodity *i* decreases with an increase in the domestic price of commodity *j*, i.e., if $\partial E_i/\partial t_j<0$ $(i\neq j)$. Note that this definition differs from the Hicksian definition as it is not based on income compensated price terms but on changes in excess demand due to changes in price distortions with real income not fixed but constrained by the production possibilities and the structure of price distortions.

---

[7] Relations (23)–(27) define the welfare index applicable to the open economy with interindustry flows. This relation has been derived and discussed by T. J. Bertrand. See also the related papers by Vanek (1970).

Noting that if complementarity between commodities does not occur ($\partial E_j/\partial t_e<0$ for all commodities) and that t_e is extreme (either $t_e>$ or $<t_j$, the distortion on all other commodities), it is immediately seen from equation (30) that community welfare improves when a) t_n is increased if $t_n<t_j$ and b) t_n is decreased if $t_n>t_j$.

This establishes Theorem 4:

THEOREM 4: *In a multi-commodity model with no complementarity, it is necessary in order to achieve a second best solution and sufficient for an increase in welfare that variable extreme distortions be eliminated.*

Since the one by one reduction of extreme distortions unambiguously improves welfare in a world with some unchangeable tariffs, the overall second best solution (with all variable tariffs optimally adjusted) must be characterized with no variable tariffs which are extreme.

Theorem 4 is the central finding of our second best analysis.[8] It provides a general guideline to policy formulation in a world where first best conditions cannot be immediately attained. These guidelines have two aspects. First, a general rule to reduce extreme distortions as far as possible—that is, until all distortions which are variable are intermediate to the highest and lowest distortions that cannot be altered. Second, if a policy of trade liberalization is to be carried out, even when all tariffs are variable, it may not be feasible to carry out all the required changes in commercial policy simultaneously. In these circumstances, policy authorities can follow the general rule of eliminating extreme distortions until the first best solution is eventually attained.

Theorem 4 is also applicable to the analysis of production taxes or subsidies which have no effect on consumers behavior. A change in the excess demand will occur solely due to the production effect. Thus, if it is assumed that an increase in the producer's price of a particular product increases the output of that good and reduces the output of all others, all variable extreme production taxes or subsidies should be eliminated, i.e., reduced or raised until they are no longer extreme.

It should be noted that Theorem 4 can remain valid even if complementarity exists between some commodities. All that is required is that the term in equation (30) for the noncomplementary commodities dominate those for the complements.

Finally, equation (30) is a generalized form of the Meade welfare index. The economic rationale of Theorem 4 is that changes in the extreme tariff will tend to reallocate expenditures towards products where the excess of the value to consumers over the real cost to the economy (defined by the fixed world prices) is greatest or reallocate resources away from industries where the excess of domestic over real costs (again defined by fixed world prices) is greatest.

III. Distortions through Taxes and Subsidies on Production

The theorems established in the preceding section are conditional on the effects of price changes on excess demands. It is also interesting to consider the problem when protection for industries is defined by the degree of distortion of value added. This is important because the profitability of different industries depends on value added or net prices rather than prices per se and sophisticated pressure groups may seek to maintain effective rather than nominal protection. We therefore consider the situation where value added distortions, brought about by taxes and subsidies imposed on the production process, are fixed for some industries and variable for others.

---

[8] Theorems of this nature have been derived by R. G. Lipsey and R. K. Lancaster for a three-commodity model, by Vanek (1964) for trade liberalization in a multi-country world, and by Hatta for welfare improvements in a single country where consumption and production distortions cause a divergence between marginal rates of transformation and substitution for different producers and consumers.

The distortions we are concerned with may be defined as the policy induced percentage change of value added from that given by the international prices. That is, where $\bar{t}_i$ is the tax ($\bar{t}_i<0$) or subsidy ($\bar{t}_i>0$) rate for individual commodities, the *effective* rate of tax or subsidy, f_i, is defined as;

$$
\begin{aligned}
f_i &= \frac{P_i(1+\bar{t}_i) - \sum_j P_j(1+\bar{t}_j)a_{ji}}{P_i - \sum_j a_{ji}P_j} - 1 \\
&= \frac{P_i\bar{t}_i - \sum_j \bar{t}_j a_{ji}}{P_i - \sum_j a_{ji}P_j}
\end{aligned}
\tag{31}
$$

where P_i is again the internationally determined price.

Since consumption distortions do not arise with taxes or subsidies in production, the welfare of the community is maximized when the consumption possibilities available to the community are maximized. The consumption possibilities are given by the budget constraint as

$$
\begin{aligned}
\sum_i P_iC_i &= \sum_i P_iX_i - \sum_i \sum_j P_ia_{ji}X_j \\
&= \sum_i \pi_iX_i
\end{aligned}
\tag{32}
$$

The gross production possibility set is

$$T(X_i, \ldots, X_n) = 0 \tag{33}$$

and the profit maximizing condition is

$$\frac{T_i}{T_j} = \frac{\pi_i(1+f_i)}{\pi_j(1+f_j)} \quad (i+1, \ldots, n\, j=1, \ldots, n) \tag{34}$$

Since the consumption distortion does not arise here due to the use of subsidies or distortion offsetting consumption taxes, the equilibrium condition for consumers is now

$$\frac{u_i}{P_i} = A \quad (i = 1, \ldots, n) \tag{35}$$

From (18), (35), and (32), we have

$$\frac{du}{A} = \sum_i P_idC_i = \sum_i \pi_idX_i \tag{36}$$

By (33), we have (37)

$$\sum_i T_idX_i = 0 \tag{37}$$

which, in conjunction with (34), yields,

$$\sum_i \pi_i(1 + f_i)dX_i = 0 \tag{38}$$

substituting (38) into (36), we have

$$
\begin{aligned}
\frac{du}{A} &= \sum_i P_idC_i = -\sum_i \pi_if_idX_i \\
&= -\pi_ef_edX_e - \sum_j \pi_jf_j\partial X_j
\end{aligned}
\tag{39}
$$

where f_e is an extreme effective rate and j denotes all other commodities. But from the budget constraint (32), we obtain

$$f_e\pi_edX_e = f_e \sum_i P_idC_i - f_e \sum_j \pi_jdX_j \tag{40}$$

Using (40) in (39), we have for a change in f_e

$$
\begin{aligned}
\frac{1}{A}\frac{du}{df_e} &= \sum_i P_idC_i \\
&= \frac{\sum_j (f_e - f_j)}{(1 + f_e)} \pi_j \frac{dX_j}{\partial f_e}
\end{aligned}
\tag{41}
$$

Since $1+f_e>0$ for produced commodities, after defining X_j as a complement in production with X_i if $\partial X_j/\partial f_i>0$, we establish a theorem closely related to Theorem 4.

THEOREM 5: *In the multi-commodity model with no complementarity in production, it is necessary in order to achieve a second best solution and sufficient for an increase in welfare that variable extreme distortions be eliminated.*

Note that Theorem 5 is directly relevant to the theory of effective tariff rates if the objective is to maximize the consumption possibilities set available to the community. Even if consumption distortions cannot be ignored or offset by consumption taxes, as this would imply, these findings are still relevant to the analysis of the production effects of effective tariffs.

REFERENCES

B. Balassa, "Tariffs, Intermediate Goods, and Domestic Protection: Comment," *Amer. Econ. Rev.*, Dec. 1970, *60*, 968–69.

T. J. Bertrand, "Welfare Indexes with Interindustry Flows: Comment," *J. Polit. Econ.*, forthcoming.

W. M. Corden, "Effective Protection Rates in a General Equilibrium Model: A Geometric Approach," *Oxford Econ. Pap.*, July 1969, *21*, 135–41.

S. E. Guisinger, "Negative Value Added and the Theory of Effective Protection," *Quart. J. Econ.*, Aug. 1969, *83*, 415–35.

T. Hatta, "A Critical Analysis of Welfare Economics," M.A. Thesis (Japanese), International Christian Univ., Tokyo 1968.

R. G. Lipsey, and R. K. Lancaster, "The General Theory of Second Best," *Rev. Econ. Stud.*, 1956, *24*, No. 1, 11–32.

V. K. Ramaswami and T. N. Srinavasan, "Tariff Structure and Resource Allocation in the Presence of Factor Substitution," in *J. Bhagwati et al., eds., Trade, Balance of Payments and Growth,* Amsterdam 1971.

R. J. Ruffin, "Tariffs, Intermediate Goods, and Domestic Protection," *Amer. Econ. Rev.*, June 1969, *59*, 261–69.

A. H. H. Tan, "Imported Material, Inter-Industry Flows, Non-Traded Goods and Optimal Trade Policies," unpublished paper, Aug. 1968.

J. Vanek, "Variable Factor Proportions and Interindustry Flows in the Theory of International Trade," *Quar. J. Econ.*, Feb. 1963, *77*, 129–42.

———, "Unilateral Trade Liberalization and Global World Income," *Quart. J. Econ.*, Feb. 1964, *78*, 139–47.

———, "The Derivation of Professor Meade's Second Best Index for the Case Where Interindustry Flows are Present," *J. Polit. Econ.*, Mar.-Apr. 1971, *79*, 345–50.

———, "Welfare Indexes with Interindustry Flows: Reply," mimeo. Dec. 1970.

The End of the North-South Wage Differential

By PHILIP R. P. COELHO AND MOHEB A. GHALI*

The failure of the market mechanism to equalize money wages between different regions in the United States has been the subject of an extensive literature.[1] In particular, the differential in money wages between the North and the South has received much of the attention as "... the classic case ... [because of the] relative severity and longevity of the wage differential, ... " (Scully, p. 757). After accounting for differences in industry mix, the differential in wages between the North and the South remains substantial. Table 1 provides a summary of some of the results reported in previous studies. Although the dates, areas covered, and the bases for comparison differ, this summary is helpful in showing the persistence of the phenomenon and to indicate the order of magnitude of the differential.

Different studies offer alternative explanations of the phenomenon and provide, either explicitly or by implication, policy prescriptions for remedial actions. By way of illustration, consider the following examples: "The North–South wage differential as it now exists is primarily noneconomic in character; .... This wage differential is likely to continue until basic decisions are made in the political and social arena which will eliminate these two factors [discrimination and inequality of the distribution of human capital] in our society" (Scully, p. 771). The wage differential arises from "... the failure of American unionism to organize more completely" (Segal, p. 154). "An alternative way to equalize interregional wage differentials ... would be to alter through education and training the skill mix of those regions which are heavily burdened with the unskilled" (Fuchs and Perlman, p. 300).

While we do not dispute the argument that investment in human capital will lead to an increase in productivity, we dispute the evidence used to justify the policy prescriptions. If welfare comparisons are to be made, they should be based on real rather than money incomes. The evidence presented in this paper strongly indicates that in spite of the observed money wage differentials, real wages, both hourly and annual, are not different in the two regions when due account is taken of differences in industry mix. Furthermore, when mix differentials are ignored, the weighted average real wage, where the weights are the number of workers employed, is *higher* in the South than it is in the North.

These results should not surprise an economist as they are the same as those predicted by supply and demand analysis. Economic theory specifies that, from the worker's viewpoint, the relevant variables are the real wage and real income.[2] Unlike incomes of other factors of production, labor earnings are primarily spent within the region in which they are earned. The cost of living differs between regions. If money wages are equalized, real wages will be different, leading to migration of labor to the region with the lower cost of living. Equilibrium in the labor market requires the equal-

---

* Assistant professors of economics, St. Mary's University, Halifax, and the University of Hawaii, Honolulu, respectively. We wish to thank the managing editor and a referee for valuable comments. Responsibility for any errors lies with the authors.

[1] The most recent contributors to this literature have been Harry Douty and Gerald Scully. The earlier literature is quite extensive; see Richard Easterlin, Frank Hanna, Victor Fuchs and Richard Perlman, Martin Segal, Lowell Gallaway, and J. W. Block.

[2] This implication of rational behavior is also supported by statements of union objectives. In reference to real wages and budget studies, the AFL-CIO has stated:

> A fundamental union goal is wages sufficient to support a family on at least a reasonable standard of living .... By indicating the minimum wages necessary for a worker to provide a reasonable living (by current American standards) for a wife and two children, the budget figure serves as a significant yardstick for many wage negotiations .... The budget figures are a helpful guide for determining the amount of wage increases which may be needed to assure for the worker's family a living level considered reasonable by the standards of their country. [as quoted in Jules Backman, p. 54]

TABLE 1—DIFFERENTIALS IN MONEY WAGES FOR PRODUCTION WORKERS IN MANUFACTURING

Author	Year Studied	Area of Coverage	Differential (Percent)	Base
Hanna[a] (p. 153)	1947	South	−11.85	U.S. = 100
Fuchs and Perlman (p. 293)	1929	South Atlantic	−14.8	U.S. = 100
	1929	East South Central	−21.2	U.S. = 100
	1947	South Atlantic	−10.8	U.S. = 100
	1947	East South Central	−16.4	U.S. = 100
	1954	South Atlantic	−10.4	U.S. = 100
	1954	East South Central	−12.6	U.S. = 100
Segal (p. 149)	1947	South	−18.7	Northeast = 100
	1954	South	−17.3	
Gallaway (p. 265)	1954	South Atlantic	−14	Mid-Atlantic = 100
	1954	East South Central	−15	
Scully (p. 768)	1958	South	−11.6	Northeast = 100

[a] This differential is the arithmetic mean of the state differentials which Hanna reported. The "South" in this case excludes Delaware and Oklahoma; in all other cases, the regions designated conform to the definition of the U.S. Census Bureau.

ity of real wages between regions, i.e., a differential in the money wage equal to the differential in the cost of living.

I. The Data and Empirical Findings

The *Census of Manufactures 1963* provides data on the number of production workers, total man-hours, and total wages for production workers in each region, state, and metropolitan area, classified by industry. The relevant data (coverage by the *Census of Manufactures* and cost of living indices) were available for ten cities in the North and the South which we chose as our sample.[3] The cost of living index for each metropolitan area was obtained from the *Handbook of Labor Statistics, 1968*. These indices measure comparative living costs based on the city worker's budget for 1966. They were deflated by the cost of living index for each metropolitan area in 1966 relative to 1963.[4] A summary of the data is provided in Table 2. While the average hourly money wage for the five northern metropolitan areas is 8.5 percent higher than that of the five southern metropolitan areas, the average hourly real wage is 1 percent higher in the South. Similarly, while the average annual wage for the northern metropolitan areas is about 5 percent higher, the average real annual wage for the southern metropolitan areas is about 4 percent higher than the average for the North.

Inference cannot be drawn, however, from comparison of these averages. They are unweighted averages, ignoring the size of each metropolitan area. More objectionable is the fact that they do not distinguish between industry and regional effects. It is conceiv-

[3] Throughout this paper, we have used the term price index to be synonymous to family budget deflator. The Department of Commerce obtains family budgets by calculating what it would cost to maintain the *same living standard* is different regions of the country; a price index shows what the *same bundle of goods* would cost in different regions of the country. Thus the Department of Commerce recognizes the same standard of living in different regions requires a different bundle of goods because of climate, for example. These cost of living indices were available for only a relatively few urban regions throughout the country. They were also available for rural areas throughout an entire region but not by state. For this reason our sample was restricted to standard metropolitan areas rather than states. However, as will be shown, the magnitude of money wage differential between North and South in our sample, 12.4 percent, is consistent with the magnitude of money wage differentials reported in the literature.

[4] The cost of living indices used are those for "family budget for city worker." The data were available in two forms: a) an index for the cost of living in each city relative to *U.S.* average for 1966, and b) an index for each city's cost of living relative to itself over time. The first of these indices was multiplied by the ratio of the index for the city in 1963 relative to 1966, to obtain our deflator.

TABLE 2—MONEY WAGES, ANNUAL WAGES, REAL WAGES AND REAL ANNUAL WAGES FOR SELECTED METROPOLITAN AREAS, 1963

Metropolitan Area	Hourly Money Wage	Annual Money Wage	Cost of Living Deflator	Hourly Real Wage	Real Annual Wage
Philadelphia	\$2.687	\$5331.4	94.3	\$2.850	\$5653.6
Pittsburgh	3.181	6178.4	91.9	3.461	6723.0
Boston	2.538	5005.2	102.9	3.466	4864.2
Buffalo	3.055	6205.5	99.1	3.083	6261.9
New York & NE New Jersey	2.497	4858.4	104.0	2.401	4671.5
Simple Average (North)	2.792	5515.7		2.852	5634.8
Washington, D.C.	2.789	5538.8	95.8	2.911	5781.6
Atlanta	2.359	4829.4	86.7	2.721	5570.2
Baltimore	2.726	5439.8	90.4	3.016	6017.4
Houston	2.777	5741.9	86.2	3.222	6661.2
Dallas	2.214	4668.3	87.6	2.527	5329.2
Simple Average (South)	2.573	5243.6		2.879	5871.9

able that North–South wage differentials are concealed by the effects of differences in industry composition.

To distinguish between the regional effect and the industry effect, the wage rate and the annual wage were calculated for every industry (two-digit classification, with SIC 21 dropped for lack of sufficient data) and for every metropolitan area.[5] It was assumed that the wage (W) in industry j in region i is represented by the following:

$$W_{ij} = \alpha + \beta_i + \gamma_i,$$

where α is the intercept, β_i is the region effect, and γ_j is the industry effect. The wage rate was regressed on nineteen dummy variables.[6] The first dummy variable assumes a value of unity for southern metropolitan areas and zero for northern ones. Each of the next eighteen dummy variables assumes a value of unity for one of the nineteen industries considered and zero for every other industry, with industry SIC 20 as a base. The first two columns of Table 3 summarize the findings for the regressions with money wages and annual money income as dependent variables. Utilizing the data on hourly wage rates, the composition of industry and the North–South dummy account for 71 percent of the observed variability. The industry dummy variables' coefficients are of no interest to the comparison of wage rates in the North and South. Rather, it is the coefficients of the South–North dummy variable that is of interest. The estimated differential, thirty-one cents, is highly significant. It implies that wages in the South are about 12.4 percent lower than the average of the entire sample. The magnitude of this differential is consistent with those reported in Table 1. As for annual earnings, the estimated North–South difference is \$528.82 and is highly significant. The annual earnings in the South are about 10.5 percent lower than the average for the entire sample.

These differentials disappear, however, when real wages are considered. The last columns of Table 3 report the results of estimating the regressions when the dependent variables are the real hourly wage and the real annual earnings. The estimated differential in real hourly wage is less than eight cents, and is insignificantly different from zero at the 10 percent level. This result im-

---

[5] No observations were available for the following industries in: Washington, SIC 22, 31, 33, 37; Baltimore, SIC 31; Houston, SIC 22, 31; Atlanta SIC 38; Dallas, SIC 31; Pittsburgh, SIC 22; Boston, SIC 24; Buffalo, SIC 31. Thus instead of having 190 observations (19 industries in each of 10 cities), we had 178 observations because 12 observations were missing. For definitions of the industries, see Table 3.

[6] This is an analysis of variance model with no interaction. Testing the relevance of the North-South classification was performed by regression analysis since we are interested in the parameter estimates as well as the analysis of variance. This procedure follows Hanna's methodology.

TABLE 3—MONEY WAGES AND INCOMES, REAL WAGES AND INCOMES: REGRESSION RESULTS[a]

	Hourly Money Wage		Annual Money Wage		Hourly Real Wage		Annual Real Wage	
Variable	Coef.	*t*	Coef.	*t*	Coef.	*t*	Coef.	*t*
Intercept	2.544		5196.6		2.578		5274.2	
N-S dummy	−.313	*−7.06*	−528.8	*−6.33*	−.076	*−1.46*	−42.6	*− .43*
Dummy of Industry								
SIC 22	−.482	*−3.33*	−1099.3	*−4.03*	−.522	*−3.07*	−1191.8	*−3.71*
23	−.619	*−4.71*	−1652.9	*−6.67*	−.659	*−4.28*	−1759.8	*−6.04*
24	−.392	*−2.91*	−1020.6	*−4.01*	−.418	*−2.64*	−1092.1	*−3.65*
25	−.233	*−1.77*	−603.9	*−2.44*	−.248	*−1.61*	−646.4	*−2.22*
26	.050	*.38*	112.1	*.45*	.064	*.42*	140.3	*.48*
27	.588	*4.48*	804.5	*3.25*	.628	*4.08*	864.9	*2.97*
28	.401	*3.06*	692.0	*2.79*	.439	*2.85*	759.2	*2.61*
29	.572	*4.35*	1307.7	*5.28*	.614	*3.99*	1393.1	*4.78*
30	−.093	*− .71*	−286.7	*−1.16*	−.096	*− .62*	−296.4	*−1.02*
31	−.631	*−3.91*	−1565.6	*−5.14*	−.637	*−3.37*	−1586.7	*−4.43*
32	.077	*.59*	212.2	*.86*	.079	*.52*	223.7	*.77*
33	.699	*5.18*	1178.9	*4.63*	.794	*5.02*	1341.2	*4.48*
34	.197	*1.50*	363.3	*1.47*	.218	*1.42*	403.1	*1.38*
35	.322	*2.45*	628.1	*2.54*	.347	*2.25*	676.2	*2.32*
36	.318	*2.42*	575.2	*2.32*	.346	*2.24*	627.4	*2.15*
37	.789	*5.85*	1786.6	*7.02*	.870	*5.50*	1961.3	*6.55*
38	.231	*1.71*	256.4	*1.01*	.250	*1.58*	278.2	*.93*
39	−.272	*−2.07*	−845.9	*−3.41*	−.290	*−1.88*	−890.1	*−3.06*
R^2	.712		.765		.650		.717	
F	20.57		27.09		15.44		21.05	

[a] The industries corresponding to the Standard Industrial Classification Code are:

SIC 20 = Food and Kindred Products
22 = Textile Mill Products
23 = Apparels
24 = Lumber and Wood
25 = Furniture and Fixtures
26 = Paper and Allied Products
27 = Printing and Publishing
28 = Chemicals
29 = Petroleum and Coal
30 = Rubber and Plastics

SIC 31 = Leather
32 = Stone and Clay
33 = Primary Metal
34 = Fabricated Metal
35 = Nonelectrical Machinery
36 = Electrical Machinery
37 = Transportation Equipments
38 = Instruments
39 = Miscellaneous

plies that, taking into account the industry effects, there exists no significant differences in the real wage rate between the North and the South. The same conclusion holds for the annual real earnings: the differential is $42.59 per year, which is not significantly different from zero at any reasonable level of significance. The findings support our proposition: real wages are equalized between the North and the South, i.e., differences in money wages between the two regions for any given industry are completely offset by the difference in the cost of living.

However, it is possible that this conclusion is incorrect. The omission of other variables which pertain to the determination of wage rates might, under certain conditions, bias the estimated coefficient of the North–South dummy variable. In order to examine this possibility, additional variables were introduced into the wage equation. These variables were: sex, color, and capital-labor ratio.

Under the assumption that the composition of the labor force and the production function remained the same for 1960 to 1963,

TABLE 4—MONEY AND REAL WAGES: REGRESSION RESULTS[a]

			North-South Dummy		Percent Female		Percent Nonwhite		Capital-Labor ($1000)		R^2
	Dep. Var.	Intercept	Coef.	*t*	Coef.	*t*	Coef.	*t*	Coef.	*t*	
Sample 1 $n=154$	W	3.068	− .350	*−6.02*	− .019	*−4.02*	− .003	*− .67*			.744
	W^*	3.247	− .073	*−1.08*	− .024	*−4.43*	− .008	*−.155*			.688
	E	6258.387	−588.946	*−5.54*	−40.083	*−4.76*	− 6.568	*− .77*			.799
	E^*	6668.305	− 30.042	*− .24*	−50.346	*−5.09*	−16.618	*−1.65*			.756
Sample 2 $n=125$	W	3.026	− .366	*−6.38*	− .021	*−4.14*	− .004	*− .82*	.013	*3.70*	.817
	W^*	3.207	− .081	*−1.22*	− .029	*−4.87*	− .007	*−1.28*	.018	*4.43*	.784
	E	6042.086	−616.449	*−5.73*	−39.481	*−4.11*	− 8.624	*−1.01*	23.186	*3.53*	.850
	E^*	6445.258	− 48.635	*− .39*	−54.734	*−4.86*	−13.554	*−1.37*	33.633	*4.41*	.825

[a] The dependent variables for the different equations are: W = hourly money wage; W^* = hourly real wage; E = annual money wages; E^* = annual real wages.

the data used was obtained from the *1960 Census of Population.* These data can be used to calculate the percentage of female and the percentage of nonwhite workers to all employed workers by two-digit industries for each of the metropolitan areas. This was done for the industries considered in Table 3, with the exception of three industries (SIC 31, 38, 39) which were dropped from lack of sufficient data, thus reducing our sample to 154 observations (sample 1).

Unfortunately, the data on capital by two-digit industries is available in the Census only at the state level. It was, therefore, assumed that the capital-labor ratios calculated for the state provides an adequate measure of the capital-labor ratios for the metropolitan areas within the state. Even then, the observations on the Washington, D.C. metropolitan area as well as 16 other observations[7] had to be dropped from the sample for lack of data, whenever the capital-labor ratio is used, reducing the number of observations to 125 (sample 2).

The results reported in Table 4 are those obtained by regressing the appropriate variable (listed in the second column) on the North–South and industry dummy variables and on some, or all, of the additional explanatory variables. To save space, the estimated coefficients of the industry dummy variables are not reported.

The parameter estimate of interest to us here is the coefficient of the North–South dummy variable. Both samples reveal a money wage and earnings differential between the North and the South of a magnitude comparable to that found in Table 3, and this differential is highly significant. However, when the real wage and real annual earnings are considered the differential ranges from 7.3 to 8.1 cents in wages and from $30 to $48 in annual earnings, depending on the sample used and the explanatory variables introduced. In all cases, however, the North–South differential in hourly real wage and annual real earnings is insignificantly different from zero at any reasonable level of significance. These findings lend strong support to the hypothesis that real wages are equalized between the North and the South.[8]

---

[7] These are: SIC 30 for Pittsburgh and Philadelphia; SIC 27, 29, 30, 33, and 35 for Atlanta; SIC 23, 24, 27 and 29 for Baltimore; SIC 30, 32 for Houston; and SIC 22, 30 and 32 for Dallas.

[8] A rather interesting result is that there seems to be more discrimination in real terms rather than in money terms. In Table 4 the *t*-values and the coefficients for sex and nonwhite discrimination increase their absolute value when measured in real terms. The percent female coefficient increases its absolute size by more than 25 percent in both samples for both hourly wages and annual earnings when measured in real terms. The percent nonwhite coefficient more than doubles in sample 1 and almost doubles in sample 2 for both hourly and annual meaures. The nonwhite variable also becomes significant at various levels of significance when measured in real terms; when this variable was measured in money terms, its *t*-value was insignificantly

To account for the effect of education, a variable measuring the median number of school years completed was also introduced. The results obtained (not shown in the tables) indicate the sign of the estimated coefficient is negative, contrary to expectations, and the coefficient is significantly different from zero at the 1 percent level. Two possible explanations of this perverse sign are: a) Sex and color discriminations take the form of higher educational requirements as well as lower wages. This explanation seems plausible when the coefficients of the sex and color variables, after adjustment, are compared to those without adjustment. b) Wage increases are based on seniority, and the number of school years completed has an upward trend, such that the correlation between wage and school years is negative. However, the coefficient of the North–South dummy remains insignificant when the dependent variable is either the real hourly or the real annual wage, and the education variable is included.

II. Conclusions

The data used in this study indicate a differential in the money wage rate of about 12.5 percent, which is consistent with the order of magnitude of the differentials reported in other studies. The North–South differential in real wage, however, was found to be insignificantly different from zero.

This finding casts doubts on the validity of regional comparisons of per capita income. In his classic study of the trends in regional incomes, Easterlin found that per capita income in the South ranged from a high of 76 percent (1840) to a low of 51 percent (1880 and 1900) of U.S. per capita income. To draw any conclusions from these findings, knowledge of the time paths of regional prices is needed. For example, a large portion of the observed income differentials in 1880 and 1900 might be a result of a divergence in regional prices during the post-Civil War deflation. Or the differential might be even larger than that observed if regional prices converged during that era.

Policy prescriptions designed to eliminate the inequity which emanates from the North–South wage differentials are redundant if the observed money wage differential is the result of an efficient economic system. Arbitrary intervention to raise the money wage in the South to northern levels will result in the misallocation of resources, as long as differences in regional prices exist. Similarly distortive are the effects of a uniform minimum wage law.

---

different from zero, in all cases, at reasonable levels of significance.

REFERENCES

J. Backman, *Wage Determination*, Princeton 1959.

J. W. Block, "Regional Wage Differentials," *Mon. Lab. Rev.*, Apr. 1948, *66*, 371–77.

H. M. Douty, "Wage Differentials: Forces and Counterforces," *Mon. Lab. Rev.*, Mar. 1968, *91*, 74–81.

R. A. Easterlin, "Regional Income Trends, 1840–1950," in S. Harris, ed., *American Economic History*, New York 1961, 525–47.

V. R. Fuchs, and R. Perlman, "Recent Trends in Southern Wage Differentials," *Rev. Econ. Statist.*, Aug. 1960, *42*, 292–300.

L. E. Gallaway, "The North-South Wage Differential," *Rev. Econ. Statist.*, Aug. 1963, *45*, 264–72.

F. A. Hanna, *State Income Differentials 1919–1954*, Durham 1959.

G. W. Scully, "Interstate Wage Differentials: A Cross Section Analysis," *Amer. Econ. Rev.*, Dec. 1969, *59*, 757–73.

M. Segal, "Regional Wage Differences in Manufacturing in the Post-War Period," *Rev. Econ. Statist.*, May 1961, *43*, 148–55.

U. S. Bureau of the Census, *Census of Manufacturer: 1958*, Washington 1961.

———, *Census of Manufacturer: 1963*, Washington 1966.

———, *Census of the United States: 1960 Population*, Washington 1963.

———, *Statistical Abstract of the United States 1968*, Washington 1968.

U. S. Department of Labor, *Handbook of Labor Statistics 1968*, Bull. No. 1600, Washington 1968.

Subsidies, Economic Lives, and Complete Resource Misallocation

By P. TAUBMAN AND R. RASCHE*

There are several accepted partial equilibrium conclusions concerning the effects of subsidies on resource allocation and prices. For perfectly competitive industries, if the demand curve is downward sloping and the supply curve is upward sloping, a subsidy will raise the quantity produced and will also lower the price paid by the purchaser, but by an amount less than the subsidy. If, however, the supply curve is perfectly elastic, the quantity produced will increase and the effective price will decline by the amount of the subsidy.[1] If the supply curve is totally inelastic, the quantity produced and the effective price will remain unchanged. Finally, if the demand curve is totally inelastic, the quantity produced will remain unchanged and the effective price will decline by the amount of any subsidy.

Within a general equilibrium context, the basic conclusion is that if the initial situation were Pareto optimal, any change in the pattern of use of resources induced by a subsidy is a misallocation. The quantity of misallocation is given by the change in resources (along a given production possibility curve), and the cost to the economy per unit of the misallocated resources is equal to the difference in value of the resources in their improper and proper uses. In some circumstances, as demonstrated by Arnold Harberger, it is possible to use the partial equilibrium results to approximate closely the cost of the misallocation in a general equilibrium model.

These conclusions have been reached either for consumer goods or for capital goods whose economic life is fixed. However, the literature on the subject has not considered the effect of subsidies on asset lives. We demonstrate in this paper that if economic lives of capital goods are not fixed, the above conclusions are incorrect in two important ways. First it will be shown in a partial equilibrium model that even if the demand curve is totally inelastic, the price of capital services to the purchaser will not in general decline by the amount of the subsidy. Second, in the general equilibrium model an additional type of resource misallocation can occur, since economic lives can be shortened and consequently the stock of capital can be reduced. This type of misallocation is complete in that it diverts resources to uses with no value. Further, when this type of misallocation occurs, the excess burden calculations used by Harberger are inappropriate. The next section of this paper will discuss various theories of the determination of asset lives. Then a model in which lives are endogenous will be constructed and examined for several classes of subsidies.

I. Theories of Asset Life Determination

At least three different theories of variable asset lives have been developed. The oldest of these formulations, discussed by Knut Wicksell and known as the Ackerman problem, is concerned with choosing the optimum type of capital when each capital good produces services at a uniform rate over its physical life, but different capital goods have different physical lives.[2]

The second theory, developed by Harold Hotelling, analyzes the determination of the economic life of an asset whose net rental stream declines over time.[3] As the asset ages, the net rental income stream could decline because 1) the quantity of services declines; 2) the price received for the services declines either from changes in status or from the introduction of newer technologies; or 3) the cost of providing the services increases.

---

* Associate and assistant professor of economics, University of Pennsylvania, respectively. The research in this paper was partially supported by the U.S. Treasury Department. The conclusions, however, do not necessarily reflect the Treasury Department's views. We wish to thank G. Brannon for helpful comments, E. Villani and Mrs. J. Young for their research assistance, and the managing editor for correcting an error in an earlier draft.

[1] Effective price is defined as the price paid by the consumer.

[2] The physical lives are determined solely by technical and not economic factors.

[3] Net rental income is the difference between the revenues received and the associated costs (including

Economic life is defined as the age at which the net rental stream for the asset is less than the earnings which can be obtained by scrapping the asset and investing the proceeds.[4] More formally when there are no taxes this can be expressed:

$$\text{Find } N = t \text{ such that} \qquad (1)$$
$$Y_t = iX_t - \Delta X_t$$

where

Y_t = net rental income in the year t of the asset's life.
X_t = salvage value in the year t of the asset's life.
i = rate of return which can be earned by investing the proceeds of scrappage.
N = economic life of the asset.

Economic life obviously need not be as long as physical life. For example, even when the scrap value is zero, the asset could exist and still be capable of providing services, but not yield a profit.

The third formulation of the variable life problem stems from the user cost concept of J. M. Keynes. If the rate at which services are provided by an asset is subject to the owner's choice, but a greater flow of current services entails lower service capabilities in the future, then, as shown by Taubman and Maurice Wilkinson, the optimal service rate will respond to changes in such things as the purchase price of capital and the interest rate. But the greater are the services provided in the earlier years of an asset's life, the shorter will be the physical and, in appropriate circumstances, the economic life of the asset. In this paper we deal primarily with the Hotelling formulation while providing reference to the other approaches where they have relevance to our analysis.

II. Subsidies and Economic Lives

In this section we consider a model for a perfectly competitive industry whose product is capital services. For ease of exposition we assume that one group of people owns the capital while a second group rents the services.[5]

To emphasize the differences from the standard model, we shall assume that the demand curve for the asset services is completely price inelastic. Since the analysis is not affected by a general income tax, we shall also assume that the tax rate is zero.

As noted above, net rental income can deteriorate over time because of several different reasons. To postpone certain complications in measuring and aggregating capital, we begin with the simplest case in which the net rent curve declines only because the cost of operations, $C(k)$, are a monotonically increasing function of the age, k, of the asset. Thus we will assume that services yielded by the capital are constant over the asset's life and that there are no technological or status changes during the asset's lifetime.[6]

Because there are no status considerations and because old and new capital yield the same services, the rentals paid, R, will be the same for capital of all ages. Since investments, $I(t)$, of the different years yield the same service, it is possible to define a stock of capital, $S(t)$, as the sum of the services in the still existing investments. In a long-run equilibrium, all buildings will have a life of N years; hence, once N is calculated for a new investment, it is immediately known how much investment from previous years is still in existence. Thus

$$S(t) = \sum_{k=0}^{N} fI(t-k) \qquad (2)$$

where f is the flow of services per time period per unit of investment. If we further assume that the equilibrium is one of no growth and

---

repairs and maintenance to the capital) in providing the capital services.

[4] While Hotelling's initial contribution discussed real estate, this formulation is the foundation for replacement theory.

[5] This separation corresponds to the markets for commercial space, rental residential properties, etc.

[6] We also assume that this no change situation is expected to occur and to continue.

that investment has been the same in the previous T years, then

$$N\,[fI\,(t)] = S(t) \tag{3}$$

The remaining equations in the model can be written as:

$$P = \sum_{k=0}^{N} [R - C(k) + B(k)]/(1 + r)^k \tag{4}$$

$$I(t) = Q(r, i) \qquad \frac{\partial Q}{\partial r} > 0; \quad \frac{\partial Q}{\partial i} < 0 \tag{5}$$

$$R - C(N + 1) + B(N + 1) < iX \leq R - C(N) + B(N) \tag{6}$$

$$N[fI(t)] = D^* \tag{7}$$

In the above equations P is the cost of new assets, $B(k)$ is the subsidy at age k, i is the rate of return on alternative investments, and X is the scrap value of the used asset.[7] Equation (4) defines the internal rate of return, r, of the asset. This internal rate of return, together with the rate of return on other investments, determines the level of current investment as indicated by equation (5). Finally equation (6) is the condition which determines the economic life, N, and equation (7) is the equilibrium condition between supply and the inelastic demand for capital services D^*.

Since this is a partial equilibrium analysis, we assume P, i, X, and $C(k)$ are determined exogenously. Equations (4)–(7), the exogenous variables, and the assumptions made in the preceding paragraphs are sufficient to determine N, R, r, and I.

Assume first that $B(k) = 0$ for all k and denote the solution values as N^0, R^0, r^0, I^0. We wish to consider the effect on these variables of introducing subsidies, $B(k)$, with various time patterns. An exact analytic solution of the problem is difficult to obtain because of the manner in which N is determined in equation (6), and because the rate of return on investment must be evaluated over a finite and variable life. However, it is possible to determine the class of subsidies which will leave N unchanged.

Since I is assumed to be a strictly monotonic function, it follows from (7) that if $N = N^0$, then $r = r^0$ is required to maintain I^0. Under these conditions (4) is satisfied only if:

$$\sum_{k=0}^{N^0} [1 + r^0]^{-k}[R + B(k) - R^0] = 0 \tag{8}$$

For any R satisfying this condition it is necessary to show that inequalities of (6) are still satisfied at N^0. First consider the subsidy $B(k) = M$ for all k. The condition $R^1 = R^0 - M$ is necessary for (8) to hold. But now (6) becomes

$$R^1 - C(N^0 + 1) + M < iX \leq R^1 - C(N^0) + M \tag{9}$$

which on substituting reduces to

$$R^0 - C(N^0 + 1) < iX \leq R^0 - C(N^0) \tag{10}$$

Hence if R^0 is a solution with economic life N^0, R^1 is also a solution with life N^0 when $B(k) = M$ for all k.

Next suppose that the dollar value of the subsidies declines monotonically as the investment ages. The necessary condition for the solution of (8) is that

$$(R^0 - R^2) \sum_{k=0}^{N^0} \frac{1}{(1 + r^0)^k} = \sum_{k=0}^{N^0} \frac{B(k)}{(1 + r^0)^k} = \overline{B} \tag{11}$$

or

$$(R^0 - R^2) = h\overline{B}, \tag{12}$$

where

$$h = \frac{1}{\sum_{k=0}^{N^0} (1 + r^0)^{-k}} \quad \text{and} \quad \overline{B} = \sum_{k=0}^{N^0} \frac{B(k)}{(1 + r^0)^k}$$

Since the subsidies decline monotonically with age

---

[7] We assume that the scrap value is independent of the age at which the asset is scrapped. The analysis is not changed substantively if X is made a function of the age of the asset. See also our discussion of land below.

$$(13) \qquad B(0) > h\bar{B} > B(N^0)$$

$$(14) \qquad (R^0 - R^2) = h\bar{B} > B(N^0)$$

For $N = N^0$ equation (6) requires that $iX \leq R^2 - C(N^0) + B(N^0)$. As long as $C(k)$ is a continuous monotonic function, this part of (6) will be satisfied by an equality, but with the change from R^0 to R^2 this condition is not met as seen from (11) and (14).

If it is assumed that r must remain constant because i is unchanged, then it is easy to demonstrate that this class of subsidies requires $N^2 < N^0$. That is, if N were to increase, a given flow of investment would generate a larger stock; thus, with demand fixed, R must decline below R^2. With $C(k)$ and $B(k)$ monotonically increasing and decreasing respectively, this R cannot be a solution to (6); hence, N must decline.

Suppose, however, r is allowed to vary. Substitution of equations (5) into (7) yields a negatively sloped function, denoted as A, for the feasible combinations of N and r. But equation (6) guarantees that $R - C(k) + B(k)$ is nonnegative for all terms in the right-hand side of (4). Hence, if N is increased, r must also be increased to maintain the equality of (4). Call this positive sloped locus G. But in our discussions of equations (12)–(14) which are derived from (4) and (6), we showed that given any r the introduction of a monotonically decreasing subsidy will cause a reduction in the N which satisfies (4) and (6). Thus, G must shift to the left with such subsidies. When combined with A, this implies a higher r and a lower N.

In either instance the subsidy causes a misallocation. If lives are shortened, then to maintain the fixed level of capital stock more investment will be necessary. If lives are lengthened, the misallocation occurs because assets are used that would be too costly to operate in the absence of the subsidy.

In the previous analysis, net income declined as the asset aged because operating costs were a monotonically increasing function of age. Next assume that operating costs are constant but that the services yielded by an asset decline with age. In this instance we can add up the services from each existing vintage of investment to obtain a measure of the capital stock. The only difference required in the analysis used earlier is that the services from new investment be equal to the services in the investment retired. As long as the flow of services over time is exogenously determined, the previous analysis is valid.

The final case to consider is the one in which the price paid for services of an investment declines as the investment ages. Since each investment is a different vintage, we can only aggregate the vintages into a capital stock after converting into value terms using the relative prices of various vintages. This measure of capital stock will vary if relative prices change, but in a steady state relative prices of various vintages are constant. This measure of capital stock will be fixed and there will be only one measure of capital. We can also have a demand function for this aggregate measure of capital. In any steady state equilibrium we can determine the functional relationship between rents and the age of the asset. The parameters of this relationship are functions of the cross-elasticities of demand between vintages. If the economic life is shorter and the number of vintages reduced, the function may shift. However, we can assume that lives remain constant and the functional representation is unchanged, and then we can determine which subsidies are consistent with this assumption.

Let $R(k) = a + bk + ck^2$ where $R(k) =$ rent paid on vintage k and let a subsidy of $B(k) = M$ for all k be introduced. As in our previous analysis, the subsidy will reduce the rental paid for capital. Let us reduce $R(k)$ by changing the constant term to $(a - M)$ for all k. Then following our previous analysis, rents for each vintage will decline by an amount equal to M; hence, at each age the subsidy will leave net rental income unchanged. In this instance the constant subsidy will not alter N, though monotonically changing subsidies will. However if the age pattern of rents is represented by other functional forms (e.g., $R(k) = R_0 e^{-gk}$) and if we give a constant subsidy M to all vintages, net rental income will not remain unchanged for all vintages. Therefore other age patterns of subsidies would be required for lives to be unchanged.[8]

However certain rules about the changes in economic lives can be deduced, because (with demand fixed) the discounted value of the subsidy must equal the discounted value of the rent reduction. Suppose that a plot of both the subsidies and the rent reduction against age reveals that the curves coincide or intersect only once. In the first instance economic lives are unaffected while in the second case lives will decrease if the subsidy curve initially lies above the rent reduction curve. This case also leads to somewhat weaker conclusions on resource misallocation. While it is still true that the resources tied up in an investment are being destroyed too soon, the monetary value of waste is less, because by assumption the value of the services declines with age. Alternatively it might be noted that purchasers can benefit by the reduction in the age of the stock since they prefer newer to older assets.

The previous discussion was conducted in terms of the age pattern of the subsidy. It is useful to consider the real world equivalents to the various subsidies discussed.[9] It is difficult to think of any subsidy that increases after-tax income by a given dollar amount every year nor can we find an example of a subsidy which increases with the age of the asset. Even a reduction in the tax rate for one industry would cause a proportionate rather than a constant change in after-tax income. It is possible, however, to enumerate some subsidies which are concentrated in the early years of an asset's life. The investment tax credit in use from 1962–69 granted the entire subsidy into the first few years of life.[10]

The second such type of subsidy is a "too rapid" tax depreciation allowance system. A tax depreciation allowance system is too rapid if the present discounted value of the tax deductions exceeds the present discounted value of economic depreciation. While "accelerated" depreciation need not be a subsidy, we have shown elsewhere that even straight line depreciation is a subsidy for office buildings, an asset to which the above analysis is relevant. (see Taubman–Rasche (1969 a, b))[11]

Since in our model it is the tail that wags the dog, a comparison of the effects of accelerated depreciation and the investment tax credit must concentrate on the net after-tax income at the "end" of the asset's life. Assume that two subsidies are designed so that for the existing economic life, their present discounted values are equal. Since the present discounted values are equal, the new equilibrium rent levels will be the same under both subsidies. Let the investment tax credit subsidy be given entirely in the first year of the asset's life while the depreciation subsidy is spread out over a number of years. If the accelerated depreciation subsidy becomes zero for the last l years of the asset's life, the effect on economic lives will be the same as the investment tax credit. If, however, the accelerated depreciation subsidy is negative in these l years, then the accelerated depreciation will introduce shorter lives than the investment credit. Of course, the opposite happens if the subsidy is positive at the end. In our earlier empirical work, (see Taubman–Rasche (1969a)) we have found that at the end of the asset life there is a negative subsidy for all conventional tax depreciation schemes. Thus the impact of accelerated depreciation on economic lives would be larger than a tax credit.

Before looking at some general equilibrium interpretations of our results, it is useful to consider briefly a few other details. Note first that if the demand curve is not perfectly inelastic, any subsidy would have an impact on the quantity of capital demanded. However it is easy to show that the useful life effect would also occur in addition to the impact predicted by traditional theory. It is also possible to demonstrate why, when the demand curve is vertical, the Ackerman and user cost determinations of asset lives are not affected by the form of the subsidy. The Ackerman model, as given in Wicksell, is concerned with the choice of that asset whose physical life maximizes net worth. With a fixed demand curve, a subsidy of a fixed dollar amount in each year will be offset by a decline in rent by the same amount. Suppose instead that a monotonically decreasing subsidy of the same present discounted value as the constant dollar subsidy were enacted. Since the present discounted value of the rent reductions just considered was equal to the present discounted value of the constant dollar subsidy, the same rent reduction would just offset the monotonically decreasing subsidy. Since this offset is consistent with the initial optimal physical life, net worth is unchanged by the subsidy scheme and the life which maximized net worth continues to do so. A similar proof holds for the user cost case.

---

[8] In the example cited, the change in market rents declines with increasing age. If a subsidy constant at all ages were granted, net rental income for older vintages would be increased after adjustment and economic lives would be lengthened.

[9] The results of the previous analysis in the presence of a general income tax are unchanged. For a discussion of real world subsidies it is appropriate to use after tax income and rates of return.

[10] While the credit applied to investment of year t, limitations on the amount that can be used in t along with carry-forward provisions spread the credit over several years.

[11] Various long-run partial equilibria under alternative tax depreciation policies have been examined in Taubman-Rasche (1969a). When straight line depreciation is substituted for double declining balance methods, the capital stock declines by about .75 percent. Simultaneously the economic life increases by 5 percent. Thus the accelerated depreciation causes a misallocation of resources of 6 percent as measured in terms of capital stock.

It has been shown that when net income declines as an asset ages, economic lives will be affected by the age pattern of the subsidy even when the demand for the stock of capital is vertical. In order to maintain a constant stock of capital with shorter lives, it is necessary (in equilibrium) that more gross investment occur in each year. Assuming that the pre-subsidy situation was Pareto optimal, the reduction in life (the increase in gross investment) with the stock constant constitutes a waste of resources. The waste is complete in that once the resources are prematurely destroyed, they cannot earn any return.[12] If there were no distorting subsidies, these resources could have been used to produce consumption goods or capital goods, the demand for whose services is not totally inelastic. To attract this extra investment to the subsidized industry, the return on the capital must rise. This, of course, is another way of explaining why the purchasers of the capital services do not receive all the benefit of the subsidy when the demand curve is vertical and the subsidy shortens lives. Notice also that since the resources are put to a use with zero value, an uncritical application of the excess burden calculations as described in Harberger would be inappropriate.

The usual method of studying subsidies in a general equilibrium context is to represent the equilibrium as the tangency of the production possibilities frontier and a utility function. This solution will be Pareto optimal under well-known and generally assumed conditions. A subsidy is then shown to yield a situation in which different prices are relevant for producers and consumers. With different relevant prices a new equilibrium is located where the utility function intersects the frontier and both functions are tangent to their appropriate relative prices. The loss in welfare is the appropriate measure of the cost of the misallocation. Our analysis extends the cost of resource misallocation in the following way. If the subsidy lowers the useful lives of assets, then for given investment less capital will be available. Since the position of the production possibilities frontier depends on the amount of capital, the frontier will retreat as a consequence of the subsidy. The cost of the resource misallocation should be calculated in terms of the welfare at the old equilibrium points.[13]

Note that even if circumstances justify a subsidy, the choice of the pattern of the subsidy affects economic lives. From a Pareto optimality viewpoint, one economic life should be best (for a given welfare function); thus, some subsidies with their associated age patterns should not be used.

---

[12] The salvage value can be put to use but salvage value may refer to land rather than produced resources and in general would be less than the value when in use in unrestricted markets.

[13] As we have previously indicated, it is theoretically possible that some subsidies would lengthen economic lives. Under this circumstance the production possibilities frontier could still lay on an indifference curve representing a lower level of welfare than the initial equilibrium.

When the asset involved is a structure, there is one special feature deserving of mention. The salvage value will include the price of land. If the demand for structures increases because of a subsidy, we would expect the price of land to rise. However, from equation (6) it is clear that an increase in the salvage would further shorten economic lives and accentuate the welfare loss.

III. Summary and Conclusion

This article has been concerned with the effects of subsidies on the quantity of investment and capital stock when the economic lives of assets are determined by economic criteria. In each of the three theories of asset lives, the optimal life is a function of some appropriate set of relative prices. The contribution of this article is not in saying that these relative prices could be affected by subsidies since such a conclusion follows immediately from the assumptions. What we have shown is that subsidies with different age patterns (but with the same present discounted value) will have different impact on lives. Moreover, it has been demonstrated that in discussing partial and general equilibrium impacts of subsidies, it is necessary to distinguish between the stock of capital and economic lives since even if the demand for capital is completely inelastic in a particular market, only one class of subsidies will leave economic lives and resource allocation unaffected. It was also noted that a subsidy such as the investment tax credit induces a smaller change in lives than a switch from straight line to accelerated depreciation. Finally it was shown that while some subsidies shorten and others lengthen lives, the resource misallocation effects were not symmetric. The asymmetry arises because when lives are too short the assets are put to a use with a zero return; however, when lives are too long, the asset is at least used in an endeavor with a positive value.

This analysis has implications that are important for several topics. First of all it suggests that if circumstances dictate that the government grant a subsidy,[14] proper resource allocation not only requires that the present discounted value of the subsidy be correct, but that the age pattern of the subsidy be such as to give the optimal life. Second, this paper implies that if for a noticeable amount of assets, lives are not exogenous but are determined by economic factors, much of the recent empirical literature on the impacts of investment subsidies has understated the effects.[15] Third, since the costs of setting the economic life too low are greater than those from setting the life too high, and since knowledge of the optimal life is imperfect, it would seem that policy would be determined so that there is bias towards too long lives.

REFERENCES

E. H. Bowman and R. B. Fetter, *Analysis for Production Management,* Homewood 1957.

R. E. Hall and D. W. Jorgenson, "Tax Policy and Investment Behavior," *Amer. Econ. Rev.,* June 1967, *57,* 391–414.

A. Harberger, "Taxation, Resource Allocation and Welfare" in J. Due, ed., *The Role of Direct and Indirect Taxes in the Federal Revenue System,* Nat. Bur. Econ. Res. and Brookings Institution, Princeton 1964.

H. Hotelling, "General Mathematical Theory of Depreciation," *J. Amer. Statist. Ass.,* Sept. 1925, *20,* 340–53.

J. M. Keynes, *The General Theory of Employment Interest and Money,* New York 1936.

P. Taubman and R. Rasche, (1969a) "Economic and Tax Depreciation of Office Buildings," *Nat. Tax J.,* Sept. 1969, *22,* 334–45.

——— **and** ———, (1969b) "Supply and Demand for Office Buildings: The Effect of Accelerated Depreciation, Capital Gains

---

[14] For example private benefits being less than social benefits.

[15] This conclusion applies to the partial equilibrium results derived by Robert Hall and Dale Jorgenson and to the general equilibrium results derived by Taubman and Terence Wales.

Rules and Mortgage Terms," discussion paper 137, University of Pennsylvania 1969.

——— and ———, "The Income Tax and Real Estate Investment," *Tax Incentives,* Tax Institute of America, Lexington, Mass. 1971.

P. Taubman and T. Wales, "Impact of Investment Subsidies in a Neo-Classical Growth Model," *Rev. Econ. Statist.,* Aug. 1969, *51,* 287–97.

P. Taubman and M. Wilkinson, "User Costs, Capital Utilization and Investment Theory," *Int. Econ. Rev.,* June 1970, *2,* 209–15.

——— and ———, "User Cost, Output, and Unexpected Price Changes" in E. Phelps, ed., *The New Microeconomics in Employment and Inflation Theory: Price, Wage and Employment,* New York 1970.

K. Wicksell, "A Mathematical Analysis of Dr. Ackerman's Problem," in P. Newman, ed., *Readings in Mathematical Economics,* Vol. 2, Baltimore 1968.

Factor Market Imperfections, the Terms of Trade, and Welfare

By RAVEENDRA N. BATRA AND PRASANTA K. PATTANAIK*

A rich literature has recently developed concerning the implications of the interindustry factor-price differential for some "standard" theorems in the theory of international trade. Most of the authors have been concerned with the welfare and second best effects of factor market distortions: Jagdish Bhagwati, Bhagwati and V. K. Ramaswami, Bhagwati, Ramaswami, and T. N. Srinivasan, Batra and Pattanaik (1970, 1971), Batra and Gerald Scully, Albert Fishlow and Paul David, Everett Hagen, and Harry Johnson (1965, 1967).

More recent studies, however, have been concerned with the impact of factor market imperfections on income distribution, allocation of resources, the possibility of multiple equilibria, and the pattern of international trade. Murray Kemp and H. Herberg, and Bhagwati and Srinivasan have written on the shape of the transformation curve and the output response of price changes in the presence of factor market distortions. They show that outputs may respond "perversely" to price changes irrespective of whether the transformation curve is convex or concave to the origin. Stephen Magee, on the other hand, has explored the consequences of factor price differentials for, among other things, factor intensities, factor prices, and the pattern and volume of international trade. Some of the results derived by Magee are very interesting. For example, he shows that a rise in the factor price differential may actually be detrimental to the factor which is apparently benefiting from this rise; furthermore neither the Heckscher-Ohlin theorem, nor the Stolper-Samuelson theorem may be valid.[1]

In this paper we explore the consequences of a change in the interindustry wage differential for the terms of trade as well as welfare. Some very unexpected results come out. For instance, a rise in the wage differential need not always change the commodity-price ratio against the industry that was initially paying the wage differential. Rather, the change in the wage differential may have any effect on the terms of trade. Furthermore, a rise in the wage differential may actually raise social welfare, and vice versa.

I. Assumptions and the Model

We deploy throughout the analysis the standard model of trade theory, where given supplies of two primary factors produce two traded commodities, X_1, the importable commodity, and X_2, the exportable commodity, modifying it only to introduce a constant wage differential between the two sectors. The two production functions are:

$$X_1 = F_1(K_1, L_1) = L_1 f_1(k_1) \tag{1}$$

$$X_2 = F_2(K_2, L_2) = L_2 f_2(k_2) \tag{2}$$

where K_i and L_i are the capital and labor inputs and $k_i = K_i/L_i$ is the capital-labor ratio in the ith commodity. Both production functions are assumed to be homogeneous of degree one.

Denoting u_i and v_i as the marginal products of labor and capital, respectively, in the ith commodity, we write

$$v_i = \partial F_i/\partial K_i = f_i'$$

and

$$u_i = \partial F_i/\partial L_i = f_i - k_i f_i'$$

---

* University of Western Ontario and Delhi School of Economics, respectively. Batra's research was financed by a Summer Research Grant from Southern Illinois University, and Pattanaik's research was partly supported by the National Science Foundation. We wish to express our gratitude to the referee for valuable suggestions.

[1] The contributions by Kemp and Herberg, Bhagwati and Srinivasan and Magee have been recently surveyed by Ronald Jones.

Under perfect competition, the price of each factor of production is equal to the value of its marginal product. Let w_i be the wage rate, r_i the rental rate of capital in the ith industry, and p_i the price of the ith commodity. Then

$$w_i = p_i u_i = p_i(f_i - k_i f_i'),$$

and

$$r_i = p_i v_i = p_i f_i'$$

It is assumed that the rate of return on capital is the same in both industries, but wage rates differ, that is,

$$r_1 = r_2 = p_1 f_1' = p_2 f_2',$$

and

$$\alpha w_1 = w_2,$$

or

$$\alpha p_1(f_1 - k_1 f_1') = p_2(f_2 - k_2 f_2') \qquad (\alpha \neq 1)$$

Let $p = p_1/p_2$ be the price of X_1 in terms of X_2. Factor rewards can then be expressed as:

$$(3) \qquad r = p f_1' = f_2'$$

$$(4) \quad \alpha w_1 = \alpha p(f_1 - k_1 f_1') = (f_2 - k_2 f_2') = w_2$$

Without loss of generality, p is assumed to have an initial value of unity. With full employment:

$$(5) \qquad L_1 + L_2 = \bar{L},$$

and

$$(6) \quad K_1 + K_2 = L_1 k_1 + L_2 k_2 = \bar{K},$$

where the bars indicate that L and K are in fixed supply.

Let Y equal national income in terms of X_2. Then,

$$(7) \qquad Y = pX_1 + X_2$$

Let D_i be the demand for ith commodity. Then,

$$(8) \qquad D_1 = D_1(p, Y)$$

and

$$(9) \qquad D_2 = D_2(p, Y)$$

These aggregate demand functions are assumed to have the same properties as the individual demand functions. Let E_1 be the quantity imported of X_1 and E_2 be the quantity exported of X_2. Then,

$$(10) \qquad E_1 = D_1 - X_1$$

and

$$(11) \qquad E_2 = X_2 - D_2$$

Finally, the balance-of-payments equilibrium requires that

$$(12) \qquad pE_1 = E_2$$

Throughout our analysis we assume incomplete specialization and an initial situation of free trade, so that p represents the international terms of trade.

With equation (12), our specification of a "determinate" general equilibrium model of international trade with interindustry wage differential is complete. There are twelve equations in eleven endogenous variables (X_i, k_i, L_i, D_i, E_i, and Y) and four parameters ($\bar{K}$, $\bar{L}$, α, and p). However, the model is not over-determined, because by the Walras law, only one of the two demand equations is independent.

II. Absence of Monopoly Power in International Trade

We now utilize the model described in Section I to analyze the impact of a change in the degree of distortion, α, on factor intensities, k_1 and k_2, factor prices, r, w_1, and w_2, the two outputs, X_1 and X_2, and real income under the assumption that the country in question is a price taker, so that p remains constant. Differentiating the system of equations (1)–(6) totally with respect to α and remembering that p initially equals unity, we obtain,

$$(13) \quad \frac{dk_1}{d\alpha} = \frac{f_1 - k_1 f_1'}{f_1''(\alpha k_1 - k_2)}$$

$$(14) \quad \frac{dk_2}{d\alpha} = \frac{f_1 - k_1 f_1'}{f_2''(\alpha k_1 - k_2)}$$

$$(15) \quad \frac{dr}{d\alpha} = \frac{f_1 - k_1 f_1'}{\alpha k_1 - k_2}$$

$$(16)\quad \frac{dw_1}{d\alpha}=\frac{-k_1(f_1-k_1f_1')}{\alpha k_1-k_2}$$

$$(17)\quad \frac{dL_1}{d\alpha}=\frac{-dL_2}{d\alpha}$$

$$=\frac{1}{k_2-k_1}\left[L_1\frac{dk_1}{d\alpha}+L_2\frac{dk_2}{d\alpha}\right]$$

$$(18)\quad \frac{dX_1}{d\alpha}=\frac{-(f_1-k_1f_1')}{(k_1-k_2)(\alpha k_1-k_2)}$$

$$\cdot\left[\frac{L_1}{f_1''}\left\{(f_1-k_1f_1')+k_2f_1'\right\}+\frac{L_2f_1}{f_2''}\right]$$

and

$$(19)\quad \frac{dX_2}{d\alpha}=\frac{f_1-k_1f_1'}{(k_1-k_2)(\alpha k_1-k_2)}$$

$$\cdot\left[\frac{L_1f_2}{f_1''}+\frac{L_2}{f_2''}\left\{(f_2-k_2f_2')+k_1f_2'\right\}\right]$$

Before we proceed to derive results from equations (13)–(19), a few remarks concerning factor intensities in the "physical" and the "value" sense are in order. Without loss of generality, we assume, hereafter, that X_1 is capital-intensive relative to X_2, that is $k_1>k_2$. This is the factor-intensity relationship between two commodities in the physical sense, because we compare only the ratios between physical units of the factors employed in the two industries. The factor-intensity relationship may also be expressed in the value sense, whose sign, in the presence of the wage differential, may be different from that of the relationship between physical factor intensities. The factor intensity in the value sense in each industry is given by (r_iK_i/w_iL_i), so that now X_1 is capital-intensive relative to X_2, if $(r_1K_1/w_1L_1)>(r_2K_2/w_2L_2)$, or if $(r/w_2)(\alpha k_1-k_2)>0$.[2] Since α is not equal to one, it is evident that (k_1-k_2) may not have the same sign as (αk_1-k_2). From equations (13) and (14), it may be observed that both k_1 and k_2 are functions of α, which leads to the possibility of reversals in the sign of (k_1-k_2). However, Magee and Herberg, Kemp, and Magee have shown that, in the presence of incomplete specialization, the physical factor intensities cannot be reversed as a result of a change in α or p alone. Since in this section we are concerned only with effects of a change in α at constant terms of trade, this type of factor intensity reversals is ruled out by our assumption of incomplete specialization. Nonspecialized equilibria are, however, compatible with factor intensity reversals in the value sense, which means that X_1, assumed to be relatively capital-intensive in the physical sense, may actually become relatively labor-intensive in the value sense.[3] If $\alpha>1$, the sign of (αk_1-k_2) must still be positive. Now $\alpha>1$ implies that $w_2>w_1$, so that the wage differential is paid by the producers of X_2, the labor-intensive commodity (in the physical sense). In other words, when an industry pays the wage differential on its intensive factor, the factor intensities in the value sense must have the same sign as factor intensities in the physical sense. However, the same may not be the case when an industry pays the differential on its nonintensive factor, because here $\alpha<1$.

With these remarks in mind, we now proceed to state briefly some of the results obtained by Magee, if only to facilitate comparison with our results obtained in the next section. Consider equations (13)–(16) under the condition of nonreversal of the value factor intensities. Due to the fact that production functions are linearly homogeneous, $f_i''<0$. It is clear, therefore, that a fall (rise) in α raises (lowers) the capital-labor ratio in both industries, and, as a result, lowers (raises) the reward of capital and raises (lowers) the reward of labor in both industries.[4] The effects of a change in the degree of distortion (α) on the two outputs can be obtained from (18) and (19). With $f_i''<0$, the square bracketed terms in both equations are negative, so that $dX_1/d\alpha>0$ and $dX_2/d\alpha<0$. In other words, a rise in α raises the output of X_1 and lowers the output of X_2, and vice versa. Here the output response to a change in α is "normal."

---

[2] The distinction between factor intensities in the physical and the value sense is a common feature of many of the articles written recently in this area. See, for examples, Magee, Herberg et al., and Jones.

[3] The reversal of value factor intensities is determined by the relative elasticities of substitution in the two industries. For further details on this point, see Magee, and Herberg et al.

[4] It can be easily seen that w_2 and w_1 move in the same direction. Therefore, throughout the paper we determine the sign of a change in w_1 only, the effect on w_2 following the same direction.

If the value factor intensities get reversed as a result of the change in α, then all these results also get reversed, because now (k_1-k_2) and (αk_1-k_2) have opposite signs. The output response to a change in α will now be "perverse."[5]

We are now in a position to determine the impact of a change in the degree of distortion on real income or welfare. Differentiating (1) and (2) totally and dividing through by dX_1, we have:

$$\frac{dX_2}{dX_1}=\frac{u_2dL_2+v_2dK_2}{u_1dL_1+v_1dK_1} \tag{20}$$

Furthermore the commodity-price ratio, p, from (3)–(6) can be written as:

$$p=-\frac{u_2dL_2+v_2dK_2}{\alpha u_1dL_1+v_1dK_1}, \tag{21}$$

so that from (20) and (21), we can write:

$$\frac{dX_2}{dX_1}=-\beta p, \tag{22}$$

where $\beta=(\alpha u_1dL_1+v_1dK_1)/(v_1dK_1+u_1dL_1)$.[6]

It is clear that $\beta\gtreqless 1$ if $\alpha\gtreqless 1$. Differentiating (7) with respect to α, utilizing (22) and remembering that $p=1$ initially, we obtain:

$$\frac{dY}{d\alpha}=\frac{dX_1}{d\alpha}(1-\beta) \tag{23}$$

The following results can be derived.

1. If the value factor intensities are not reversed, so that $dX_1/d\alpha>0$, then

$$\frac{dY}{d\alpha}\gtrless 0 \quad \text{if } \beta\lesseqgtr 1, \quad \text{or } \alpha\lesseqgtr 1$$

In other words, a rise (decline) in α results in a rise (decline) in real income when $\alpha<1$ or $w_1>w_2$, so that the differential is paid by X_1 producers. Conversely, a rise (decline) in α leads to a decline (rise) in real income if the wage differential is paid by X_2 producers. Now when $\alpha<1$, a rise (decline) in α means a decline (rise) in the wage differential. If, on the other hand, $\alpha>1$, a rise (decline) in α signifies rise (decline) in the wage differential. Thus, we may conclude that a decline in the wage differential leads to a rise in real income and a rise in the wage differential results in a decline in real income.

2. If the value factor intensities are reversed, then, from equation (18), $dX_1/d\alpha<0$. In this case,

$$\frac{dY}{d\alpha}\lesseqgtr 0 \quad \text{if } \beta\lesseqgtr 1 \quad \text{or } \alpha\lesseqgtr 1$$

This result is just the opposite of the previous result. In this case, a decline (rise) in the wage differential leads to a decline (rise) in real income.

From these two results and the remarks made earlier in this section, the following theorem may now be derived:

THEOREM 1: *If an industry pays the differential on its intensive factor, then a rise (decline) in the factor-price differential lowers (raises) the level of real national income at constant terms of trade. This conclusion is valid even if the higher priced factor to an industry is its nonintensive factor, provided the industry is intensive in the use of a factor in both the physical and the value sense. Otherwise, the factor-price differential and the level of real income are positively related.*

This theorem has important implications for a recent theorem by Bhagwati p. 23, that, for a small country, a decline in the degree of factor market imperfection necessarily raises welfare, and conversely, barring

---

[5] Following other authors in this area, we define the output response to be normal if the output of a commodity rises (declines) as a result of either a rise (decline) in its relative price for a given α, or a decline (rise) in the wage differential paid by it for a given p. The output response is defined to be perverse if the opposite holds.

[6] This proof is obtained from Hagen. Also see Batra and Scully.

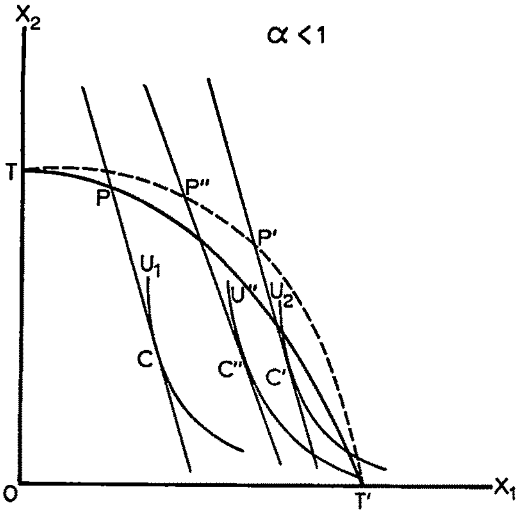


FIGURE 1

the possibility of inferiority in social consumption. We have just shown that Bhagwati's theorem does not hold if the value factor intensities are reversed as a result of the wage differential.[7]

Results 1 and 2 can be further illustrated in terms of diagrams. Consider Figure 1 where TT' is the transformation curve, P the production point and C the consumption point on the international-price ratio given by the slope of PC, with the level of welfare indicated by the social indifference curve U_1. As is well known, the effects of the wage differential are to shrink the transformation curve towards the origin and disrupt the equality between the commodity-price ratio and the marginal rate of transformation (also see equation (22)). The original transformation curve is omitted for the sake of clarity. The wage differential is shown to be unfavorable to X_1, as may be clear from the fact that the relative price of X_1 exceeds the marginal rate of transformation at all points of production. A decline in the wage differential or the factor market imperfection has the effect of an outward extension of the transformation curve to $TP'T'$. The new production point is given by P', which shows a higher output of X_1 and a lower output for X_2, as is the case when the value factor intensities are not reversed. The consumption point shifts to C' and welfare improves to U_2. Note that the terms of trade remain unchanged (PC is parallel to $P'C'$).

Figure 2 depicts the case where the value factor intensities are reversed. The output of X_1 declines and that of X_2 rises as a result of the decline in the wage differential. The new production and consumption points are again given by P' and C', but welfare declines from U_2 to U_1.[8]

III. Monopoly Power in Trade

Until now we have assumed that the country under analysis is a price taker and has no influence on world prices. This has been a convenient assumption and has enabled us to derive the results by differentiating the equations in our model while keeping p constant. However, if the country in question is large and possesses monopoly power in trade, then p can no longer be kept constant. The results derived in Section II may then have to be modified.

Our immediate task then is to find out the impact of a change in the degree of distortion on the terms of trade. The method of our analysis here is first to determine at constant terms of trade the sign of the change in the excess demand for the importable commodity, E_1. Depending on this sign, the terms of trade will eventually improve or deteriorate in order to restore equilibrium in the foreign trade market. Equation (12)

---

[7] If the value factor intensities are reversed, there will be multiple equilibria in the economy (see Bhagwati and Srinivasan). Therefore, in fairness to Bhagwati, it may be stated that his therorem holds if we rule out the possibility of multiple equilibria. However, the only restriction that Bhagwati places on his theorem requires the absence of inferior goods, but, as we have shown, it also requires the absence of multiple equilibria.

[8] Figure 1 depicts the case of a normal output response along with a concave (to the origin) transformation curve, whereas Figure 2 describes the case of a perverse output response. It may be pointed out here that neither the output response nor our results are influenced by the shape of the transformation curve. In associating the normal output response with the concave transformation curve, we are merely following the tradition set by Hagen, Bhagwati and Ramaswami, Johnson, and Batra and Pattanaik (1970), who assume a normal output response with respect to relative price changes and depict it under a concave transformation curve.

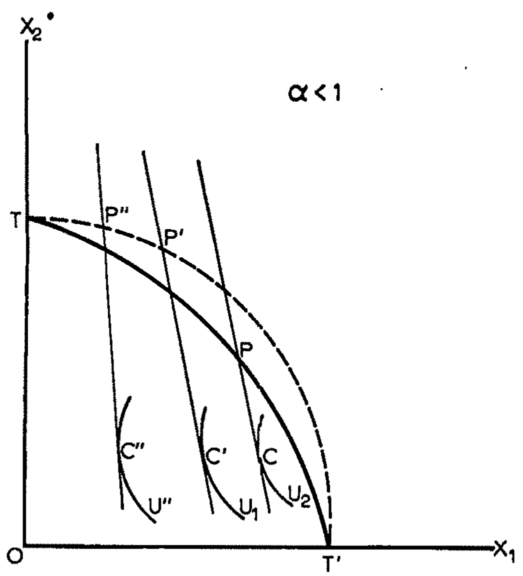


FIGURE 2

may be rewritten as:[9]

$$pE_1(p,\alpha) - E_2(p) = 0 \tag{12}$$

Differentiating (12) totally with respect to α, we have:

$$p\frac{\partial E_1}{\partial\alpha} = E_1\left[\frac{p}{E_2}\frac{dE_2}{dp} - \frac{p}{E_1}\frac{\partial E_1}{\partial p} - 1\right]\frac{dp}{d\alpha}, \tag{24}$$

so that with $p=1$ initially,

$$\frac{dp}{d\alpha} = \frac{\partial E_1/\partial\alpha}{(a_1+a_2-1)E_1}, \tag{25}$$

where $a_1 = -(p/E_1)(\partial E_1/\partial p)$ and $a_2 = (p/E_2)(dE_2/dp)$ are the price elasticities of demand for the domestic and foreign country imports.[10] Stability in the foreign trade market, which we assume throughout the analysis, requires that[11] $a_1+a_2>1$.

Now from the definition of imports, $E_1 = D_1(p, Y) - X_1(p, \alpha)$, so that

$$\frac{\partial E_1}{\partial\alpha} = \frac{\partial D_1}{\partial Y}\cdot\frac{\partial Y}{\partial\alpha} - \frac{\partial X_1}{\partial\alpha} \tag{26}$$

Let $m_i = \partial D_i/\partial Y$ be the marginal propensity to consume the ith commodity: $(0<m_i<1)$. Substituting (26) in (25), we derive:

$$\frac{dp}{d\alpha} = \frac{(\partial Y/\partial\alpha)\left[m_1 - \dfrac{1}{1-\beta}\right]}{(a_1+a_2-1)E_1} \tag{27}$$

We can derive the following additional results:

3. A change in α always changes the terms of trade. This is a trivial result and is evident from the fact that $m_1 \neq 1/(1-\beta)$.

4. Let $dX_1/d\alpha>0$, a case representing the nonreversal of the value factor intensities in spite of the wage differential. If $\beta<1$, then $m_1<1/(1-\beta)$ and, from (23), $\partial Y/\partial\alpha>0$, so that $dp/d\alpha<0$. If, however, $\beta>1$, then $m_1>1/(1-\beta)$ and, from (23), $\partial Y/\partial\alpha<0$, so that here again $dp/d\alpha<0$. Thus we may conclude that a rise (decline) in α leads to a decline (rise) in the price of the importable commodity, X_1, relative to the price of the exportable commodity, X_2, or, what is the same thing, to an improvement (deterioration) in the terms of trade. When $\alpha<1$, then $w_1>w_2$, so that a rise in α (or a decline in the wage differential paid by the importable commodity) results in an improvement in the terms of trade. The opposite holds when $\alpha<1$, but there is a rise in the wage differential. If $\alpha>1$, then $w_2>w_1$, so that a rise in α or in the wage differential paid by the exportable commodity results in an improvement in the terms of trade, and vice versa.

5. If the value factor intensities are reversed, so that $dX_1/d\alpha<0$ then result 4 is also reversed and we obtain certain unexpected results: 1) a decline (rise) in the wage differential paid by the importable commodity results in a deterioration (improvement) in the terms of trade, and 2) a rise (decline) in the wage differential paid by the exportable commodity leads to a deterioration (improvement) in the terms of trade. The

---

[9] Since E_2 is the foreign country's demand for imports, it is a function only of p but not of α.

[10] In the presence of the wage differential, outputs, as shown by Bhagwati and Srinivasan and Kemp and Herberg, may respond perversely to changes in relative prices, in which case it is possible that $a_1>0$. However, we rule out this possibility.

[11] This is the so-called Marshal-Lerner condition for stability.

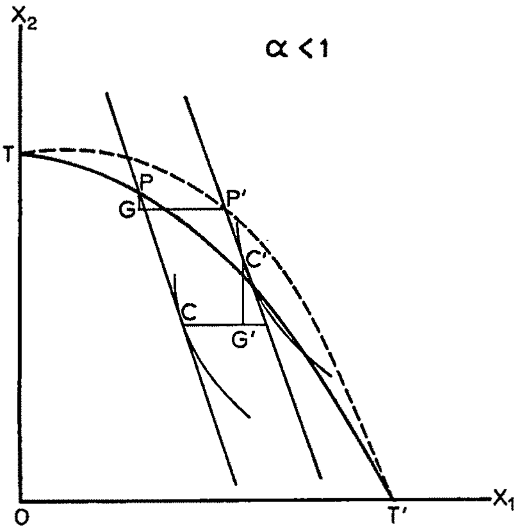


FIGURE 3

result is surprising, because the relative price of a commodity suffering from the wage differential actually declines, even though the wage differential becomes more unfavorable to it.

These results can be further clarified with the help of simple diagrams. Consider Figure 3 where α is shown to be less than one. A rise in α or a decline in the wage differential shifts the transformation curve TT' outward to $TP'T'$, production shifts from P to P', consumption from C to C' and the excess demand for X_1 declines at constant terms of trade (PC is parallel to $P'C'$), because the rise in the production of X_1, equal to GP', exceeds the rise in the demand for X_1, equal to CG'. Eventually the terms of trade will improve. Figure 4, where the output of X_1 is negatively related to α, indicating a reversal of the value factor intensities, however, shows that E_1 rises as a result of a rise in α, because here the decline in the output of X_1, equal to GP, exceeds the decline in its demand, equal to $C'G'$.

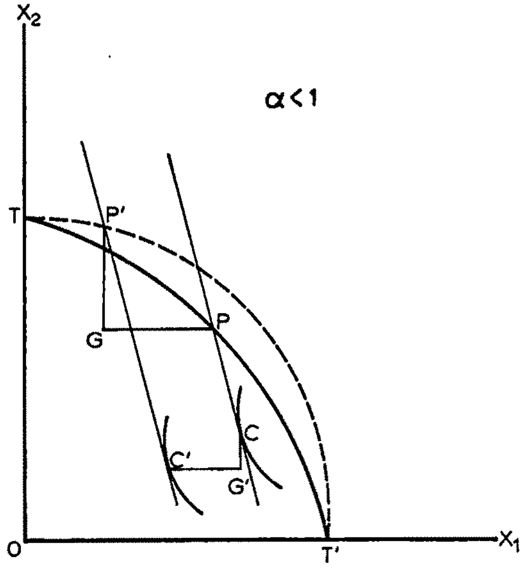


FIGURE 4

Results 3, 4, and 5 have been derived under the assumption that the import-competing good is a normal good, so that $0 < m_1 < 1$. However, if the importable is a luxury good, so that $m_1 > 1$, the results derived under the case where $\beta < 1$ may not hold.[12] Here m_1 may equal $1/(1-\beta)$, in which case the terms of trade will be unchanged in spite of a change in α. Furthermore, m_1 may exceed $1/(1-\beta)$, in which case a decline (rise) in the wage-differential paid by the import-competing industry may actually lead to a deterioration (improvement) in the terms of trade even when the value factor intensities are not reversed. On the other hand, in spite of a reversal in the value factor intensities, a decline (rise) in the wage differential paid by the import-competing industry may result in an improvement (deterioration) in the terms of trade.

It is noteworthy, however, that the possibility of $m_1 > 1$ does not affect any of the results when $\beta > 1$, because here $m_1 > 1/(1-\beta)$, so long as m_1 is positive. The story is different when the importable is an inferior good, so that $m_1 < 0$. In this case, the three results may not hold only when $\beta > 1$. Thus we conclude that if the importable is not a normal good, results 3, 4, and 5 may not hold. In the rest of the analysis, however, we abstract from this case and assume that $0 < m_1 < 1$.

---

[12] We are grateful to the referee for suggesting this possibility.

Welfare and the Wage Differential

We now turn to the effect of a change in the degree of distortion on welfare when the terms of trade also change. Since Y depends on α and p, we write

$$Y = Y(\alpha, p), \tag{28}$$

so that

$$\frac{dY}{d\alpha} = \frac{\partial Y}{\partial \alpha} + \frac{\partial Y}{\partial p}\frac{dp}{d\alpha} \tag{29}$$

Differentiating (7) with respect to p and utilizing (22), we have:[13]

$$\frac{\partial Y}{\partial p} = X_1[1 + \eta_1(1 - \beta)], \tag{30}$$

where $\eta_1 = (p/X_1)(\partial X_1/\partial p)$ is the output elasticity of the import-competing good with respect to the terms of trade. We assume that if the value factor intensities are not reversed, η_1 is positive; otherwise it is negative.[14] Substitution of (30) in (29) yields:

$$\frac{dY}{d\alpha} = \frac{\partial Y}{\partial \alpha} + X_1[1 + \eta_1(1 - \beta)]\frac{dp}{d\alpha} \tag{31}$$

From (31) we obtain the following results:

6. If the value factor intensities are not reversed, then the rise in the real income as a result of a decline in the wage differential is less than the rise in real income at constant terms of trade. Similarly, a decline in the real income as a result of a rise in the wage differential is less than the decline in the real income at constant terms of trade. These results can be explained as follows: Here $\eta_1 > 0$, and $dp/d\alpha$ from result 4 is negative, so that the second term on the right-hand side of equation (31) is negative. If $\alpha < 1$, $\beta < 1$, then from result 1, $\partial Y/\partial \alpha > 0$. Therefore, $dY/d\alpha < \partial Y/\partial \alpha$. In other words, a rise (decline) in α or a decline (rise) in the wage differential leads to a rise (decline) in the real income which is less than the rise (decline) in the real income that would result if the terms of trade were unchanged. If, on the other hand, $\alpha > 1$, $\beta > 1$, and from result (1), $\partial Y/\partial \alpha < 0$, so that here

$$\left|\frac{dY}{d\alpha}\right| < \left|\frac{\partial Y}{\partial \alpha}\right|$$

In other words, a decline (rise) in α or in the wage differential, leads to a lower rise (decline) in the real income than was the case with constant terms of trade. For geometrical explanation of these results, we revert to Figure 1, where at constant terms of trade welfare rises from U_1 to U_2 as a result of a decline in the wage differential. A decline in the wage differential when $\alpha < 1$ results in an improvement in the terms of trade and the international price line shifts from PC (parallel to $P'C'$) to $P''C''$. Production shifts to P'', consumption to C'' and welfare declines from U_2 to U''. This result is rather surprising, because an improvement in the terms of trade consequent upon a change in α, which should normally raise welfare even beyond the welfare rise at constant terms of trade, actually leads to a lower welfare. However, it can be easily explained in terms of another result by Batra-Pattanaik (1970), according to which an improvement in the terms of trade may actually lower welfare if the wage differential is unfavorable to the importable commodity, i.e., if $\alpha < 1$.[15]

7. Equation (31) clearly leads to the paradoxical possibility that a decline in the wage differential may actually lower welfare in the presence of variable terms of trade even if the value factor intensities do not get re-

---

[13] For a more detailed proof, see Batra and Scully and Batra and Pattanaik (1970).

[14] Magee and Herberg et al. have shown that if both α and p change, the reversal of physical factor intensities is compatible with incomplete specialization. The output response then becomes perverse only if factor intensities get reversed in either the physical or the value sense. If factor intensities are reversed in the physical as well as the value sense, the output response still remains normal. In order to maintain consistency with the previous analysis, we assume, hereafter, that the physical factor intensities never get reversed. The interested reader can, of course, derive different results by relaxing this assumption.

[15] This result, however requires that $\beta < 1 + 1/\eta_1$, a necessary and sufficient condition for the Batra-Pattanaik (1970) result to hold. Also see Batra and Scully.

$$\frac{dk_1}{d\alpha} = \frac{(dp/d\alpha)[k_2 f_1' + \alpha(f_1 - k_1 f_1')] + (f_1 - k_1 f_1')}{f_1''(\alpha k_1 - k_2)} \tag{32}$$

$$\frac{dk_2}{d\alpha} = \frac{\alpha f_1(dp/d\alpha) + (f_1 - k_1 f_1')}{f_2''(\alpha k_1 - k_2)} \tag{33}$$

$$\frac{dr}{d\alpha} = \frac{\alpha f_1(dp/d\alpha) + (f_1 - k_1 f_1')}{\alpha k_1 - k_2} \tag{34}$$

$$\frac{dw_1}{d\alpha} = \frac{-[k_2 f_1(dp/d\alpha) + k_1(f_1 - k_1 f_1')]}{\alpha k_1 - k_2} \tag{35}$$

versed. This result can also be derived from Figure 1.[16]

8. If the value factor intensities are reversed, result 6 is also reversed. In this case, $dp/d\alpha > 0$ and $X_1[1+\eta_1(1-\beta)]$ can also be shown to be positive.[17] Therefore it follows from (31) that $|dY/d\alpha| > |\partial Y/\partial\alpha|$. In other words, a decline (rise) in the wage differential leads to a greater decline (rise) in the real income than was the case when terms of trade were kept constant. Going back to Figure 2 gives us a better understanding of this result, for we find that it is the perverse response of outputs to relative price changes that is responsible for this result. It may be noted that the implications of such a perverse response do not come out clearly from equation (31). In Figure 2, a decline in the wage differential (or a rise in α), leads to a deterioration in the terms of trade, so that PC is replaced by a steeper line $P''C''$. The production point shifts to P'', which shows a still lower output for X_1, consumption point to C'', and welfare declines further to U''.[18]

Wage Differential, Factor Intensities, and Factor Prices

We now turn to the effect of a change in α on capital-labor ratios and factor rewards when the terms of trade also change. Differentiating (3) and (4) with respect to α, and taking $p=1$, we get equations (32)–(35).

The following results may now be derived:

9. If the value factor intensities are not reversed, then $dp/d\alpha$ from result (4) is negative. With $dp/d\alpha < 0$, there may no longer be a unique relationship between the capital-labor ratio and α on the one hand, and between factor rewards and α on the other even if $(\alpha k_1 - k_2)$ never changes sign. Specifically, a decline (rise) in α may actually lower (raise) the capital-labor ratio in each industry.

This result makes interesting comparison with the results derived in Section II under the assumption of constant terms of trade. There α and k_1 and k_2 were shown to have a unique relationship, whereas result (9) shows that, in the presence of variable terms of trade, the relationship between α and k_i may not only be not unique, but it may also be reversed.

---

[16] This possibility does not require the presence of multiple equilibria and can be easily explained in terms of another theorem by Bhagwati. According to Bhagwati, if there is more than one distortion in the system, then a decline in one of the distortions may not be welfare increasing, p. 23. In this paper, Bhagwati gives a very general connotation to the concept of a distortion and defines it as a departure from any condition of Pareto optimality. Therefore, if a country possesses monopoly power in trade, Pareto optimality for domestic welfare requires the imposition of an optimum tariff. Pursuit of a policy of free trade, which we assume throughout the paper, then represents a distortion. In other words, in the presence of the wage differential, a large country following free trade policy suffers from two distortions. That is why, as shown above, a decline in the wage differential may be welfare decreasing.

[17] Here again the condition $\beta < 1 + 1/\eta_1$ is relevant.

[18] All the diagrams have been drawn for the case where $\alpha < 1$. However, the diagrammatic analysis can be easily extended to the case where $\alpha > 1$. The major difference in this case is that the relative price of X_1 will be less than the marginal rate of transformation at all points of production.

10. From (34) and (35), it is evident that the relationship between α and factor rewards also may no longer be unique. Furthermore, a decline (rise) in α, may actually raise (lower) the rate of return on capital and lower (raise) the wage rate. This result is the opposite of the result obtained in the presence of constant terms of trade.

11. Of greater interest, however, is the case where the value factor intensities are reversed. Here $dp/d\alpha$ from result (5) is positive. With $dp/d\alpha > 0$ it is evident from (32) and (33) that α and k_i still have a unique relationship as was the case with constant terms of trade. In other words, the implications of a change in α for k_1 and k_2 remain the same when the value factor intensities are reversed, irrespective of whether the terms of trade are constant or variable.

Similarly, the relationship between α and factor rewards remains unique even in the presence of variable terms of trade when the value factor intensities are reversed.

REFERENCES

R. Batra and P. K. Pattanaik, "Domestic Distortions and the Gains from Trade," *Econ. J.*, Sept. 1970, *80*, 638–49.

——— and ———, "Factor Market Imperfections and Gains from Trade," *Oxford Econ. Pap.*, July 1971, *23* 182–88.

R. Batra and G. W. Scully, "The Theory of Wage Differentials: Welfare and Immiserizing Growth," *J. Int. Econ.*, May 1971, *1*, 241–47.

J. N. Bhagwati, "The Generalized Theory of Distortions and Welfare," working paper, No. 39, M.I.T., May 1969.

——— and V. K. Ramaswami, "Domestic Distortions, Tariffs and the Theory of Optimum Subsidy," *J. Polit. Econ.*, Feb. 1963, *71*, 44–50.

———, ———, and T. N. Srinivasan, "Domestic Distortions, Tariffs and the Theory of Optimum Subsidy: Some Further Results," *J. Polit. Econ.*, Dec. 1969, *77*, 1005–10.

J. N. Bhagwati and T. N. Srinivasan, "The Theory of Wage Differentials: Production Response and Factor Price Equalization," *J. Int. Econ.*, Feb. 1971, *1*, 19–36.

A. Fishlow and P. David, "Optimal Resource Allocation in an Imperfect Market Setting," *J. Polit. Econ.*, Dec. 1961, *69*, 529–46.

E. Hagen, "An Economic Justification of Protectionism," *Quart. J. Econ.*, Nov. 1958, *72*, 496–514.

H. Herberg, M. C. Kemp, and S. P. Magee, "The Factor Market Distortions, The Reversal of Relative Factor Intensities, and the Relation Between Product Prices and Equilibrium Outputs," mimeo. 1970.

H. G. Johnson, "Optimal Trade Intervention in the Presence of Domestic Distortions," in R. E. Caves et al., eds., *Trade, Growth and the Balance of Payments*, Amsterdam 1965.

———, "The Possibility of Income Losses from Increased Efficiency or Factor Accumulation in the Presence of Tariffs," *Econ. J.*, Mar. 1967, *77*, 151–54.

R. W. Jones, "Distortions in Factor Markets and the General Equilibrium Model of Production," *J. Polit. Econ.*, May/June, 1971, *79*, 437–59.

M. C. Kemp and H. Herberg, "Factor Market Distortions, the Shape of the Locus of Competitive Outputs, and the Relation Between Product Prices and Equilibrum Outputs," discussion paper, Univ. Southampton, Aug. 1969.

S. P. Magee, "Factor Market Distortions, Production, Distribution and the Pure Theory of International Trade," *Quart. J. Econ.*, Nov. 1971.

Population Density in Patterns of Trade and Development

By DONALD B. KEESING AND DONALD R. SHERK*

Hollis Chenery and Lance Taylor abandon Chenery's earlier position that developing countries approximate a unified set of development patterns depending on population and per capita income (1960). Separating countries with less than 15 million people into two subsets based on their trade and output patterns, they find statistically significant differences between equations fitted to the two groups. This finding leads them to conclude that at least among small countries, natural resource differences constitute a major influence. Since in their view "there is no single criterion for classifying countries according to resource endowments that is statistically feasible and theoretically satisfactory" (p. 396), they settle for an exercise in descriptive taxonomy based on equations fitted to three subsets of developing countries. The operational significance of this exercise is left obscure.

Meanwhile, in a closely related study, one of the present authors worried that what appear (in Chenery-style equations) to be country-size effects unfavorable to small countries "could be merely the reverse side of a comparative advantage for small countries in goods and services based on natural resources" (Keesing, p. 455). This problem persists in Bela Balassa's extension of the analysis.

Here we demonstrate a very simple yet surprisingly effective way of taking natural resources into account in Chenery-style regression equations. The method is to put population density to work as an explanatory variable alongside population and per capita income. Population density serves as an admittedly imperfect proxy for the availability of natural resources per head.[1]

Our equations and others like them could be interpreted along the lines of Chenery's (1960) "normal patterns," but we have serious reservations about the operational significance of such "patterns." Here we limit our investigation to equations "explaining" countries' exports and imports of manufactured goods. Our main concern is to explore the significance of density and country-size effects in world trade.

I. Regression Method and Hypotheses

Our empirical procedure is to employ population density as an additional explanatory variable, in combination with per capita *GDP* and population, in log-linear, cross-national regressions roughly similarly to those pioneered by Chenery (1960). Following Balassa, the dependent variables are limited to major trade magnitudes such as exports of manufactures as a proportion of total exports (X_m/X).

The basic form of the equations is:

$$log\begin{pmatrix}\text{dependent}\\ \text{variable}\end{pmatrix} = log\ a + b_1\ log\ N + b_2\ log\frac{Y}{N} + b_3\ log\ D + u$$

where N = population, $Y = GDP$, and D = population density (which could also be written N/A where A = area), for each of a set of countries.[2] We run separate least

---

* Associate professors of economics at Stanford University and Simmons College, respectively. We gratefully acknowledge computational assistance from Christine Branson, Bo Carlsson, and Patricia DeCoster, together with extremely useful suggestions from George Borts, Franklyn D. Holzman, and Ronald McKinnon.

[1] More complex indexes of natural resources could also be suggested—Irma Adelman and Cynthia Morris include one such index in their factor analysis. But population density appeals because it is so simple and universal, and is not based on sectoral patterns of trade or output.

[2] The presence of N in all three explanatory variables does not necessarily bias the results, but Benton Massell has written in a personal communication to Keesing " . . . as the three explanatory variables are all in log form, the equation can equivalently be written

$$log\begin{pmatrix}\text{dependent}\\ \text{variable}\end{pmatrix} = log\ a + c_1\ log\ N + c_2\ log\ Y + c_3\ log\ A + u$$

where $c_1 = b_1 + b_3 - b_2$, $c_2 = b_2$ and $c_3 = -b_3$. The two squares regressions for developing countries and for industrial countries.

Densely populated countries are expected to demonstrate a systematic comparative advantage in manufactures as against primary products, signalled for example by a significantly positive coefficient for D in equations for X_m/X. The chief reason is that in any large sample of countries, land area per head should be positively (and its inverse, population density, negatively) highly correlated with the combined availability of mineral wealth, arable land, and other economically useful endowments from nature. Hence, densely populated countries could be expected to specialize in manufactures, and lightly populated countries in primary products, along lines suggested by a land-labor version of the Heckscher-Ohlin theory. A second reason is that countries with dense populations might be expected—as a matter of statistical likelihood—to enjoy advantages in manufacturing compared to lightly settled countries in terms of costs of (and access to) transportation, communications, distribution systems, and the like.[3]

We further hypothesize that with density included as a separate explanatory variable, the coefficient for population (N) will be significantly positive (in relation to X_m/X) at the same time. This would help to show that large countries enjoy a comparative advantage in manufactures, compared to small countries, even after taking into account differences in natural resources.

II. Underlying Effects

There would be at least five classes of effects that we would expect to see reflected in the results of our equations:

1) *Density effects*: other things equal, a higher D shifts comparative advantage from resource-intensive exports toward manufactures, for reasons already suggested.

2) *Country-size effects*: other things equal, in a world with international obstacles to trade, a larger N shifts comparative advantage from other goods toward manufactured goods strongly affected by scale economies at the level of the national industrial complex.

3) *Larger-income effects*: other things equal, more Y/N means a larger home market, and thus helps shift comparative advantage toward income-elastic goods strongly affected by scale economies.

4) *Development effects*: other things equal, the higher is a country's degree of development (roughly reflected by Y/N), the greater will be its comparative advantage in complex, skill-intensive manufactures. Up to a modest level of development comparative advantage will grow, but after that it will decline in simple standardized manufactures.

5) *Openness effects*: Other things equal, countries with less N (and possibly more D) trade more of their output internationally, leading to greater exports and imports per head.

To the extent that Y/N serves as an unbiased proxy for real income and development, elasticities for Y/N in our regressions presumably reflect a combination of development and large-income effects. Elasticities for N and D pick up country-size and density effects, respectively, mingled with openness effects wherever the dependent variable involves trade per capita.

The main necessary qualification is that Y/N computed in the usual way—in local prices converted into dollars at official exchange rates—is not an accurate proxy: Two countries with equal Y/N (computed in this manner) may really be quite different in their levels of real per capita income and development.[4] Travelers know that an equal

---

formulations are identical mathematically and statistically . . . " In this light, the results appear open to alternative interpretations.

[3] We do not know how to distinguish among these two sets of influences, nor can we exclude others, notably the possibility that densely populated countries may typically follow more outward-looking policies than lightly populated countries.

[4] By different real incomes in this context, we simply refer to the results of measuring both countries' incomes at some reasonable common set of prices. Differences in development allude either to measurements using some reasonable index of per capita changes associated with modernization, or to per capita value added by local nationals in producing some set of "modern" goods and services by suitably modern methods, valuing output in both countries at the same prices.

number of *U.S.* dollars will buy much more, converted at official exchange rates, in some countries than in others. Local prices converted into dollars tend to be low, for example (other things equal) in a country that has a superabundance of underemployed labor, imposes few quantitative restrictions on imports, has low tariffs, maintains a undervalued or equilibrium exchange rate, and is seeking to overcome a scarcity of foreign exchange by promoting exports and tourism. Under opposite conditions local prices (converted at official rates) tend to be much higher, so that a given money value of output represents fewer goods.[5] If these international price differences are systematically linked to other variables (such as D and N), this could affect our regression results in hidden ways.

As an example, suppose that at any given level of Y/N, we find developing countries with high and low population densities; but whenever Y/N is the same, the densely populated countries always have lower local prices (converted into dollars) for most goods, and—by the same token—higher levels of real income and development than the lightly populated countries.[6] In this case, at each level of Y/N the more densely populated countries would presumably tend to have a more "advanced" structure of industry and trade (e.g., greater manufactured exports) than the lightly populated countries. As a result, our elasticities for D, with Y/N held constant, would be influenced not only by density effects, but also by larger-income and development effects.

In the example, the hidden relationship would presumably occur in the first place mainly because population density systematically affects trade, leading to repercussions on the local price level (converted to dollars) through effects on the balance of payments, the exchange rate, the money supply, and the country's trade policies. To this extent, while hidden relationships may affect the elasticity coefficients in our regressions, statistically significant t-ratios for D can reasonably be accepted as evidence that trade is significantly influenced by density effects, and significant t-ratios for N constitute strong evidence of country-size effects.

III. Regression Results

Table 1 summarizes our main findings for twenty-six developing countries.[7] The first

---

[5] As an indicator of real income and development, Y/N may be all the more misleading because over the middle stages of the development process, equilibrium internal prices tend to be forced down relative to external prices as a result of foreign exchange scarcities (and resulting import constraints). This decline in internal prices and costs eventually helps cause a breakthrough in exporting diversified manufactures, after which the internal price level (converted to dollars) is able to rise, much as when primary exports were first expanded. But while the external bottleneck is most severe, Y/N may stay much the same for long periods even though major gains are being made in real income and development. This is because real gains are offset by a rising equilibrium exchange rate, and hence declining internal prices (converted into dollars). Two countries with equal Y/N may stand at opposite ends of this drawn-out process.

[6] We believe that there is a systematic tendency in this direction—that is, local prices (converted into dollars) tend to be considerably lower in densely populated countries (most of which are in Asia) than in more lightly populated countries (which are mainly in Africa and Latin America). This belief is supported by our theoretical expectations, namely, that a dense population should lead to low prices through greater pressures of underemployed labor, a scarcity of foreign exchange (and money supply) from exports of natural resource products, and incentives for outward-looking policies (such as realistic or undervalued exchange rates, low tariffs and few import restrictions) in order to promote exports of manufactured goods and tourist services. Conversely, a light population should lead to high prices through abundant exports, labor scarcity, and inward-looking industrialization policies.

In the case of N differences we have no equivalent belief one way or another. Cost-price influences of country size would seem to be quite mixed. For example, country-size advantages presumably improve the trade balance and lower costs, but large countries probably raise their internal prices (converted into dollars) by following more inward-looking policies than small countries do.

[7] The twenty-six developing economies included Argentina, Brazil, Cambodia, Ceylon, Chile, Colombia, Congo (Kinshasa), Ghana, Hong Kong, India, Iran, Israel, Jordan, South Korea, Libya, Laos, Mexico, Nigeria, Pakistan, Philippines, Senegal, Taiwan, Thailand, Tunisia, United Arab Republic, and South Vietnam. In Table 1, equation (4), Chile was omitted for lack of data. Published U.N. trade data for 1965 were used in each case. For these countries we find a correlation of $-.36$ between the logs of Y/N and N; $-.19$ be-

Table 1—Regression Equations for Twenty-Six Developing Countries

	Dependent Variable	Coefficients of Independent Variables			Constant Term	$\bar{R}^2$	Durbin-Watson
		Y/N	N	D			
1.	X_m/N	2.42 (3.2)	0.83 (2.0)	1.08 (3.3)	−25.0 (3.7)	.42	1.50
2.	X_m/X	1.23 (2.3)	0.82 (2.8)	0.70 (3.0)	−19.9 (4.2)	.41	1.43
3.	M_m/N	0.91 (7.2)	−0.37 (5.4)	0.15 (2.7)	−12.2 (1.1)	.83	1.69
4.	M_{tc}/N	0.09 (0.3)	−0.09 (5.7)	0.26 (1.9)	2.6 (0.9)	.60	1.39
5.	M_m/M	0.04 (0.8)	0.02 (0.7)	−0.03 (1.4)	3.9 (8.7)	.00	2.35
6.	X_7/N	1.55 (1.7)	0.38 (0.8)	0.38 (1.0)	−18.4 (2.3)	.03	2.11
7.	M_7/N	1.03 (6.8)	−0.25 (3.0)	0.07 (1.1)	−1.2 (0.9)	.75	2.02
8.	X_7/X_m	3.00 (3.2)	0.33 (0.6)	0.96 (2.3)	−27.0 (3.2)	.30	2.02
9.	M_7/M_m	0.11 (1.8)	0.13 (3.7)	−0.08 (2.7)	2.1 (3.8)	.40	2.36

Symbols:
Y = gross domestic product;
D = population density;
M = total imports;
X_7, M_7 = exports and imports of machinery and transport equipment (SITC 7);
M_{tc} = imports of textiles and clothing (SITC 65, 84, 85).
N = population;
X = total exports;
X_m, M_m = exports and imports of manufactured goods (SITC 5–8);
All variables are expressed in logs; t-values shown in parentheses, where 2.1 is significant at 5 percent level.

equation shows that per capita exports of manufactures are positively and strongly related to population density as well as per-capita *GDP*. The elasticity for N is also positive and almost significant despite openness effects. Equation (2) (the key one of the series) demonstrates even more conclusively that both population and its density are positively linked to the successful exportation of manufactures as opposed to primary products.

Equation (3) reveals an asymmetrical relationship on the side of imports. Other things equal, densely populated countries import more manufactures than lightly populated countries. Equation (4) shows that this result is not caused simply by imports of complex producer goods, since it extends to imports of textiles and clothing.[8] Equation (5) shows that intercountry varia-

---

tween the logs of Y/N and D; and .19 between the logs of N and D. Among the twenty-two industrial countries listed in fn. 11, these correlations are, respectively, −.08, −.39, and .46.

[8] Reasons for this relationship are not fully clear. Densely populated countries may be more open to imports as a matter of transport costs and/or policy. Perhaps these imports (including intermediate textile products) reflect larger-income effects in disguise plus requirements for producer goods, associated with greater specialization in manufacturing. In any case, this finding casts a rather different light on the corresponding export picture.

TABLE 2—REGRESSION EQUATIONS FOR TWENTY-TWO INDUSTRIAL COUNTRIES

	Dependent Variable	Coefficients of Independent Variables			Constant Term	$\bar{R}^2$	Durbin-Watson
		Y/N	N	D			
1.	X_m/N	1.67 (3.9)	0.03 (0.2)	0.44 (3.0)	−9.8 (2.8)	.45	2.17
2.	X_m/X	0.24 (0.8)	0.34 (3.0)	0.26 (2.4)	−6.8 (2.7)	.54	2.20
3.	M_m/N	1.00 (5.1)	−0.46 (6.4)	0.12 (1.7)	1.5 (1.0)	.76	2.85
4.	M_{te}/N	1.61 (5.8)	−0.61 (6.0)	0.20 (2.0)	−11.1 (4.9)	.77	2.69
5.	M_m/M	−0.10 (0.6)	−0.01 (0.2)	−0.04 (0.7)	0.6 (0.4)	.00	1.82
6.	X_7/N	2.46 (4.9)	0.22 (1.2)	0.57 (3.3)	−19.2 (4.7)	.59	1.91
7.	M_7/N	0.61 (1.9)	−0.54 (4.6)	−0.02 (0.2)	4.8 (1.8)	.60	2.25
8.	X_7/X_m	0.79 (4.4)	0.19 (2.9)	0.12 (1.9)	−9.4 (6.4)	.60	1.56
9.	M_7/M_m	−0.39 (1.6)	−0.09 (0.9)	−0.14 (1.6)	3.3 (1.6)	.15	2.29

Symbols: see Table 1.

tions in the composition of imports, as between manufactures and primary products, are not well explained by the three variables under consideration.

The final four equations involve exports and imports of machinery and transport equipment (SITC category 7), an important category of manufactures that we would expect to be strongly affected by economies of scale and, by extension, country size. The export equations are rather unrevealing, no doubt because developing countries have their most severe comparative disadvantage in these products—as a group they imported \$11.5 billion (at f.o.b. prices) in 1966 and exported only \$0.5 billion, even counting substantial reexports.[9] But the import picture is interesting, especially equation (9), which shows that, other things equal, countries with larger populations import more of these products relative to other manufactured imports. Presumably this is because large countries are more successful in import substitution across the whole spectrum of manufacturing and are thus able to concentrate their imports on capital goods.[10]

Table 2 presents analogous regression results for twenty-two industrial countries.[11]

---

[9] See United Nations *Statistical Yearbook* (1968, Table 15).

[10] From other sources we have reason to believe that import substitution in capital goods is most successful in the largest developing countries. For example, Chenery's (1960) normal patterns show output per head of machinery and transport equipment to be three times larger (at Y/N of \$300) when N is 50 million than when it is 2 million (Table 8).

[11] The countries were Australia, Austria, Belgium (including Luxembourg), Canada, Denmark, Finland, France, Greece, Iceland, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, United States, and West Germany. Results have been tested for sensitivity to the

In the export equations and the contrasts between the export and import equations we again find clear evidence of country-size and density effects. Except for the equations involving machinery and transport equipment, results are broadly similar to the patterns found in Table 1 among developing countries.

The strength of both country-size and income effects is particularly impressive in the case of machinery and transport equipment. Clearly the comparative disadvantage of small countries in manufactured goods in general, and in capital goods in particular, is *not* "merely the reverse side of a comparative advantage . . . based on natural resources."

We should add that in almost all these equations, the addition of D served to improve the statistical fit compared to using Y/N and N alone.

IV. Conclusions

Population density turns out to be an important influence on patterns of trade and development. As expected, a densely settled population has its largest impact in terms of increased exports of manufactured goods relative to primary products.

This relationship suggests that only the most densely populated small developing countries can look forward to early successful export specialization in manufacturing industries other than those based on proximity to natural resources. A possible further implication is that exports of industrial products well suited for production in developing countries would tend to remain concentrated in a few countries regardless of trade liberalization.

It remains unclear to what extent density effects are caused by influences favorable to densely populated countries, or only by negative influences of missing resources. Perhaps for lack of natural resources countries are simply forced to lower their dollar prices and wages until exports of some other kind become feasible. But possibly a country lacking exportable natural resources benefits from compensatory mechanisms that help to even up its chances for modernization in a world in which migration is largely blocked: exports of manufactures may lead to favorable scale effects or other dynamic advantages.

Country-size effects coexist with density effects. With natural resources taken into account by putting density into the regression equations, there is still powerful evidence of country-size effects. This is the strongest evidence to date that large countries enjoy major advantages over small countries in manufacturing. These country-size effects appear most striking in the case of capital goods. In combination these findings confirm, link, and in some ways go beyond the main previous findings of Chenery and Taylor, Keesing, and Balassa.

REFERENCES

I. Adelman and C. T. Morris, *Society, Politics, and Economic Development: A Quantitative Approach*, Baltimore 1967.

B. Balassa, "Country Size and Trade Patterns: Comment," *Amer. Econ. Rev.*, Mar. 1969, *59*, 201–04.

H. B. Chenery, "Land: The Effects of Resources on Economic Growth," in K. Berrill, ed., *Economic Development*, London 1964.

———, "Patterns of Industrial Growth," *Amer. Econ. Rev.*, Sept. 1960, *50*, 624–54.

——— and L. Taylor, "Development Patterns: Among Countries and Over Time," *Rev. Econ. Statist.*, Nov. 1968, *50*, 391–416.

D. B. Keesing, "Population and Industrial Development: Some Evidence from Trade Patterns," *Amer. Econ. Rev.*, June 1968, *58*, 448–55.

D. Usher, *The Price Mechanism and the Meaning of National Statistics*, Oxford 1968.

United Nations, *Commodity Trade Statistics*, New York 1965.

———, *Yearbook of International Trade Statistics*, New York 1966.

———, *Statistical Yearbook*, New York 1966, 1967, 1968.

---

inclusion of Greece, Iceland, Portugal and Spain. Excluding these four countries reduces the fit of the equations slightly but has little effect on the elasticity coefficients.

The Elasticity of Expansion and Scale Adjustments of the Firm

By DALE B. TRUETT*

It is surprising, given the predominance of the two-factor model of production and the pedagogical usefulness of the two-factor approach, that scale adjustments in the isoquant diagram are so little understood. Some light was shed on the foregoing when Albert M. Levenson and Babette Solon, John Rowe, and F. W. McElroy (1967) discussed the links between isoquant spacing, the expansion path, and returns to scale in this *Review*, and when Lowell Bassett and McElroy (1970) related returns to scale to long-run cost in the *Southern Economic Journal*. However, these writers dealt with expansion of the firm at constant input prices, and their analyses are applicable only to the shape of the long-run average cost curve (*LAC*) and not to its shift in response to changing input prices. A recent article by C. E. Ferguson and Thomas Saving in this *Review* has provided much needed information regarding the shift of *LAC* when a factor price change occurs. It is the purpose of the present paper to relate the Ferguson-Saving analysis to the graphics of the two-factor case and to introduce a new concept, the *elasticity of expansion*, which will prove useful in the analysis of scale adjustments in two-factor models of production.

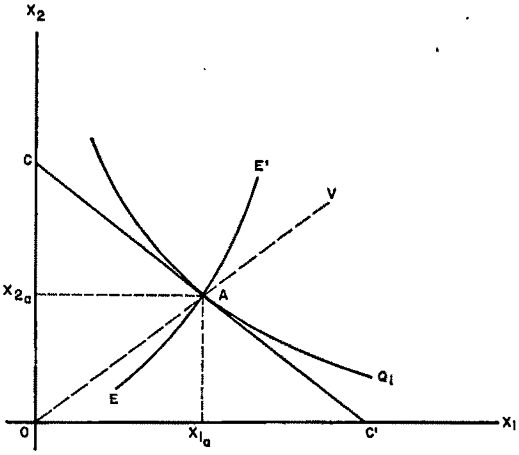


FIGURE 1

Ferguson and Saving have shown that for the perfectly competitive firm minimum *LAC* varies directly with the price of any single input, but that optimum long-run output of the firm may remain constant, increase, or decrease in response to a change in an input price, depending upon the expenditure elasticity of the price-changing input. For the two-factor case, the production function is $Q=f(x_1, x_2)$. With given input prices, p_1 and p_2, denote η_1 as the expenditure elasticity of input x_1. Then:

$$\eta_1 = \frac{c}{x_1}\frac{dx_1}{dc} \tag{1}$$

where c is the total outlay of the firm. If $\eta_1 > 0$, factor x_1 is classified as superior, neutral, or normal with respect to expenditure as $\eta_1 \gtreqless 1$; it is inferior if $\eta_1 < 0$. Ferguson and Saving have demonstrated that long-run optimum output varies directly with the price of a normal or inferior factor and inversely with the price of a superior factor.[1]

In Figure 1, assume that point A corresponds to the level of expenditure and output consistent with the minimum point of the *LAC* (the production function is nonhomogeneous and nonhomothetic). For convenience, assume $p_2 = 1$. Then C is the total outlay consistent with least-cost output when the input price ratio is $p_1/p_2 = OC/OC'$. Now denote ϵ as the *elasticity of expansion*.

---

* Chairman, department of economics, Florida International University. The author was assistant professor of economics at the University of Florida when this note was written. Thanks are due an anonymous referee for helpful suggestions on an earlier draft of the paper.

[1] For a proof of this proposition, see Ferguson and Saving.

$$\epsilon = \frac{\dfrac{dx_2}{dx_1}}{\dfrac{x_2}{x_1}} \tag{2}$$

With constant input prices, the elasticity of expansion measures the response in utilization of one of the inputs to a given change in utilization of the other input *when long-run cost-minimizing conditions are fulfilled.* In reference to Figure 1, the numerator of (2) is the slope of the expansion path EE',[2] whereas the denominator is the slope of vector OV. In what follows, it will be shown that the expenditure elasticities of inputs x_1 and x_2 can be determined from the elasticity of expansion, and, therefore, that the horizontal shift in LAC accompanying a given input price change can be determined from the relationship between the slope of the expansion path and the factor ratio (slope of OV) at a least-cost point such as A.

Consider the expenditure elasticity of x_1, η_1 as defined in (1) above. Since

$$c = x_2 + x_1p_1 \tag{3}$$

when $p_2 = 1$,

$$\frac{dc}{dx_1} = \frac{dx_2}{dx_1} + p_1 \tag{4}$$

Further:

$$\frac{c}{x_1} = \frac{x_2 + x_1p_1}{x_1} = \frac{x_2}{x_1} + p_1 \tag{5}$$

Substituting (4) and (5) into (1), one obtains

$$\eta_1 = \frac{c}{x_1}\frac{dx_1}{dc} = \frac{\dfrac{x_2}{x_1} + p_1}{\dfrac{dx_2}{dx_1} + p_1} \tag{6}$$

Note that the first term in the numerator of (6) is the factor ratio and that the first term in the denominator is the slope of the expansion path. Thus, the elasticity of expansion is intimately related to the expenditure elasticity of factors. In particular, it can be seen from (2) and (6) that if:

$$\begin{aligned} &0 < \epsilon < 1, \quad \eta_1 > 1 \\ &\epsilon = 1, \quad \eta_1 = 1 \\ &\epsilon > 1, \quad 0 < \eta_1 < 1 \end{aligned} \tag{7}$$

Returning to Figure 1, it follows from the above that when the slope of the expansion path exceeds the factor ratio (EE' is steeper than OV), $0 < \eta_1 < 1$. Thus, for the case depicted where Q_1 corresponds to initial optimum long-run output, x_1 is normal and an increase in p_1 would shift minimum LAC upward and to the right. Further, it is intuitively obvious that if x_1 is normal, x_2 *must be* superior. For, at constant input prices, if the proportion of total outlay expended upon x_1 falls, that expended upon x_2 must rise. Accordingly, it may be stated without further proof that if:[3]

$$\begin{aligned} &0 < \epsilon < 1, \quad 0 < \eta_2 < 1 \\ &\epsilon = 1, \quad \eta_2 = 1 \\ &\epsilon > 1, \quad \eta_2 > 1 \end{aligned} \tag{8}$$

From Figure 1, one would conclude that $\eta_2 > 1$ and that an increase in p_2 would shift LAC upward and to the left.

To summarize, for positively sloped expansion paths, dx_2/dx_1 and ϵ will be positive, and both expenditure elasticities will be positive. If the expansion path cuts the

---

[2] It is necessary to keep in mind that dx_2/dx_1 is the slope of the *expansion path* and not the slope of an isoquant (technical rate of substitution), since attention is directed to the behavior of factor proportions as output *changes* in the long run. In equation (4), dx_2/dx_1 is similarly defined, since factor quantities are allowed to adjust while long-run output *changes* and cost is minimized. A general expression for the slope of the expansion path has been provided in Rowe. From the usual conditions for output maximization subject to an expenditure constraint, Rowe shows that:

$$\frac{dx_2}{dx_1} = \frac{f_1f_{12} - f_2f_{11}}{f_2f_{12} - f_1f_{22}}$$

[3] The expenditure elasticity of x_2 is:

$$\eta_2 = \frac{\dfrac{1}{p_1} + \dfrac{x_1}{x_2}}{\dfrac{1}{p_1} + \dfrac{dx_1}{dx_2}}$$

factor-ratio vector from below, as in Figure 1, x_2 will be superior and x_1 will be normal. If the expansion path were tangent to the factor-ratio vector at A, or if the expansion path coincided with the vector, both factors would be neutral. Finally, if the expansion path were to cut the factor-ratio vector from above, x_1 would be superior and x_2 normal. With this information, for each case the response of optimum output to an input price change can be determined in accordance with the Ferguson-Saving analysis.

The usual assumption of divisibility of inputs suggests that a factor can only become inferior at some level of output or expenditure after having been noninferior (superior, neutral, or normal) over some lower range of output or expenditure. In such a case, the expansion path will turn back on itself as in Figure 2. The slope of the expansion path will become negative in the range where either input is inferior; thus the elasticity of expansion will be negative. In Figure 2, for example, $\eta_1<0$ for outputs greater than Q_2 (input prices given), since the amount of x_1 falls as total outlay increases. The additive term in the rightmost expression of (6) above suggests that η_1 can be positive when dx_2/dx_1 and ϵ are negative. This would require that $p_1>|dx_2/dx_1|$, but examination of Figure 2 will show that the foregoing cannot occur if isoquants are convex to the origin.[4] Thus, when

$$\epsilon < 0, \qquad (9)$$
$$\text{either } \eta_1 < 0$$
$$\text{or } \quad \eta_2 < 0,$$

depending upon whether the expansion path traverses upward to the left or downward to the right, or, alternatively, upon whether $dx_2/dc \gtrless 0$. In such a case, the expenditure elasticity of the noninferior input must be greater than one, since it is clear that the proportion of the budget expended upon it will increase. Returning to Figure 2, if point D were the point of minimum LAC for the price ratio OC/OC', one would conclude that $\epsilon<0$ and $\eta_1<0$ but $\eta_2>1$. In this case, an increase in p_1 would cause minimum LAC to shift upward and to the right, while an increase in p_2 would shift minimum LAC upward and to the left. If x_2 were inferior, the expansion path would turn downward and to the right rather than upward and to the left, ϵ would be negative, and the response to input price changes would be the reverse (in terms of optimum output) of that in the case of Figure 2.

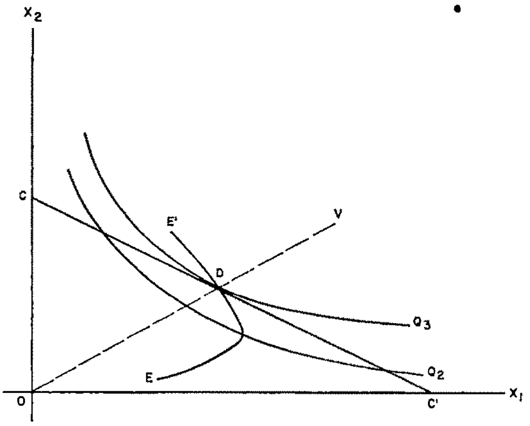


FIGURE 2

In conclusion, for the two-factor case the expenditure elasticity of inputs and the output response of minimum LAC to an input price change can be determined from ϵ, the elasticity of expansion. In the isoquant diagram, ϵ is the ratio of the slope of the expansion path to that of a factor-ratio vector constructed through any point on the path. To determine the nature of the adjustment in long-run optimum output accompanying an input price change, one need only construct a factor-ratio vector through the expansion path point corresponding to minimum LAC. If the slope of the vector and that of the expansion path are the same, $\epsilon=1$ and both factors are of unitary elasticity so that a change in the price of either will not change optimum long-run output. If the slope of the expansion path is positive but differs from that of the vector (i.e., if the two intersect as in Figure 1), one of the two factors will be superior to the other, and optimum output will change when an input price change occurs. In particular, optimum long-run output will vary directly with the price of the normal input and inversely with the price of the superior input. Finally, if the expansion path has a negative slope, one of the factors is inferior, and optimum output will vary directly with the price of that factor but inversely with the price of the other factor.

---

[4] Because of the earlier numeraire assumption ($p_2=1$), p_1 is the factor price ratio or the absolute value of the slope of the isocost line. Thus, the expansion path would have to cut isocost lines at more than one point, which is not possible under the usual convexity assumption.

In the long run, if a perfectly competitive industry is one of decreasing or increasing cost, pecuniary external economies and diseconomies will affect the firm's input prices. Unless the production function of the firm is homothetic, these externalities can be expected to lead to a change in the firm's optimum long-run output. For the two-factor case, the direction of change in optimum output depends upon the elasticity of expansion, so that from ϵ it can be determined whether efficient producing units will be characterized by larger or smaller long-run economic capacity when specific adjustments in factor prices are expected.

REFERENCES

L. Bassett, "Returns to Scale and Cost Curves," *Southern Econ. J.*, Oct. 1969, *36*, 189–90.

C. E. Ferguson and T. R. Saving, "Long-Run Scale Adjustments of a Perfectly Competitive Firm and Industry," *Amer. Econ. Rev.*, Dec. 1969, *59*, 774–83.

A. M. Levenson and B. Solon, "Returns to Scale and the Spacing of Isoquants," *Amer. Econ. Rev.*, June 1966, *56*, 501–05.

F. W. McElroy, "Returns to Scale and Cost Curves: Comment," *Southern Econ. J.*, Oct. 1970, *37*, 227–28.

———, "Returns to Scale and Spacing of Isoquants: Comment," *Amer. Econ. Rev.*, Mar. 1967, *57*, 223–24.

J. W. Rowe, "Returns to Scale and the Spacing of Isoquants: Comment," *Amer. Econ. Rev.*, June 1968, *58*, 548–50.

Monopsony in the Market for Public School Teachers

By John H. Landon and Robert N. Baird*

The effect of monopsony in the labor market on wage levels is an empirical question. The most convenient place to test the usual micro-theory hypothesis is within a regional industry where skills and nonmonetary job aspects are similar but the degree of regional market concentration varies widely.[1] The test is most convincing when the skill level of the job is high and cross elasticity of supply between this and other occupations is low.[2] This paper attempts to demonstrate the effect of monopsony in the market for public school teachers.

Several recent publications have reviewed economic aspects of the determination of teachers' salaries. For example, Don Hellriegel, Wendell French, and Richard B. Peterson conducted a behavioral analysis of collective bargaining attitudes among teachers and found, among other things, that lower relative salaries are associated with positive attitudes toward collective negotiations. However, they did not expand their study to see if collectively negotiated salaries were significantly higher than nonnegotiated salaries. Two more recent works have explored this matter. Hirschel Kasper studied average teachers' salaries and percentages of teachers unionized, both by state, and found no consistent relationship between the two. Robert Thornton, however, used data for eighty-seven large school systems and found that unionization was significantly and positively related to teachers' salaries. Both of these studies omitted an analysis of the degree of competition in those labor markets. Werner Hirsch studied the market for public school teachers in California. While he did not specifically derive supply and demand schedules for teachers he argued that legal and quasi-legal obstacle in that state have prevented that labo market from attaining equilibrium, resulting in "wage control shortages" for severa specializations. Joseph Kershaw and Roland McKean showed how school district adherence to salary schedules that ignored the economics of supply and demand not only created shortages of teachers among educational specializations sought by the private sector, but also impeded the ability of schoo districts to make dramatic improvements in the quality of instruction.

In this paper, we seek to add to this growing body of knowledge by investigating the effect of monopsony power of local schoo districts on the determination of teachers salaries. The basic hypothesis behind this test is that competition in the market for teachers should result in higher salary levels The exact formulation of this test and the results are presented below. We conclude that local school districts do compete strongly with each other for first-year teachers. As a result, salaries of teachers are lower where districts are few and large in a given metropolitan area, and salaries are higher where many small districts compete for teaching personnel. This conclusion seems to us to imply that taxpayers may stand to gain (lose) and teachers may stand to lose (gain), in the event of a major move to consolidate (decentralize) urban school districts.

I. Determinants of Teachers' Salaries

The principal determinants of teacher salaries must be considered if the effect of monopsony is to be correctly isolated. Three such determinants are 1) the ability of the district to pay; 2) the willingness of the district to pay; and 3) the elasticity of the supply of teachers to the school district.[3]

---

[*] Assistant and associate professors, Case Western Reserve University. We wish to thank the referee for several helpful comments.

[1] For a theoretical discussion of this question, see, among others, Joan Robinson, pp. 211–28.

[2] John Landon has completed such a test for the newspaper industry.

[3] It is useful here to distinguish between the elasticity of supply to a geographic region and the elasticity of the

Each of these factors is briefly discussed below.

The ability of a community to pay for education depends on both the extent of local resources and the availability of state and federal funds for this purpose. The greater the per student financial resources of the community, the less sacrifice is involved in providing any given level of educational quality. This ability to pay may be measured in terms of the value of taxable property in the community and/or the level of income per capita or per student.

The value placed by the community on the public education of its children should be the key to its willingness to pay the bill for a quality system of elementary and secondary schools. This willingness will interact with ability to pay in determining the demand for teachers, administrators, and educational capital in the school district. The minimum willingness to pay may be mandated by the state in the form of teacher certification requirements, pupil/teacher ratios, or per capita spending requirements. Maximum willingness is limited only by the aspirations of the community. Willingness is reflected in such measures as the ratio of taxes to income or assets.

The availability of alternative employment within the teaching profession at the local level strongly influences the responsiveness of the labor supply to a given change in compensation. This is the principal focus of this paper. The presence of other school districts in the area increases the number of options available to the teacher without requiring a change of residence. In a metropolitan area with a dozen or more school districts, a teacher with strong local ties has reasonable alternatives without leaving the area or the teaching profession. Just as a large number of school districts increases the options of present teachers, it also offers incoming educators more alternatives. When several school systems within an area are expanding, those teachers with the highest qualifications will presumably locate in the districts offering the highest levels of compensation, other things being equal. Some competition among the districts in starting salary should be induced. In order to keep its present teachers or add well qualified teachers to its staff, a district must remain competitive with surrounding districts.

A heavily populated area with a single public school system makes that system in a very real sense a monopsonist. Teachers currently in the system can change employment within the profession only at the expense of moving to another city. Only the more distant substitutes of nonteaching employment and the private schools offer real alternatives. Therefore, in the absence of growth, the system may for long periods pay salaries well below those of comparable districts in other areas. Even in a growing system, it is possible that the natural inflow of teachers whose spouses find employment in the area may provide adequate staffing in the absence of competitive salaries. Only rapid growth in the absence of such an inflow of secondary wage earners would force the system to adopt a more attractive salary structure.

From the foregoing reasoning, we anticipate that competition for teachers within metropolitan areas is an important determinant of salary levels. The only influence likely to offset this intraregional competition is collusion among local districts. To the extent that they agree on standard rates for incoming teachers, the districts in the area could create an effective monopsony. The probability of an effective buying cartel should decline as the number of districts increases and as the market for teachers becomes tight.

This analysis would not be complete, however, without recognizing an additional factor bearing on the extent of monopsony: the comparability of the districts. If for example, one district is far superior in terms of facilities, students, or in some other respect, substitutability is decreased. However, as long as wage rates are not uniform

---

supply within that region. The elasticity of supply *to* a geographic region involves such considerations as teacher certification requirements, population migration patterns, and (in the long run) the region's system of higher education. In this paper we are concerned only with supply elasticity *within* the region.

and change over time, there is no particular reason to believe that substitution will not take place at the margin.

II. Regression Analysis

Multiple regression analysis of 1966–67 national cross-sectional data on 136 school districts is employed below to isolate the effect of monopsony on the level of teachers' salaries. As the dependent variable we use the district's contractual salary for the beginning teacher with no prior experience.[4] As the principal independent variable we selected the logarithm of the number of school districts in the county containing that district. The rationale is, of course, that monopsony power decreases as the number of school districts in the county increases. By using logarithms, we imply that variations in the beginning salary are positively correlated with percentage variations in the number of districts. We realize that county boundaries will not always be coterminous with labor market boundaries, but no suitable alternative is available.

In selecting the remaining independent variables, we have attempted to find suitable measures of the most obvious factors that would enter into a reduced form equation. As measures of ability to pay, we include per capita income in the county and percentage of total school revenues from local sources. Since per capita income reports by school district are not available, we use county data, even though intracounty variation in incomes may be extensive, especially among central cities and the more remote suburban districts. The percentage of revenue raised locally is also a measure of ability to pay since equalization programs within states seek to aid poorer districts relatively more than the richer ones. The coefficients for these variables are expected to be positive. As a measure of willingness to pay, we use effective property tax rates by district. Theoretically, individuals who favor support for education are willing to pay higher taxes for a better school system. Accordingly, the regression coefficient for this variable should be positive.[5]

Ordinary least squares (*OLS*) estimation was applied to the data described above, with the results of the regressions presented in Table 1.[6] In equation 1, all 136 districts are included. The signs of the coefficients all have the expected values, and the coefficient of the monopsony variable is significantly different from zero at the .005 level.

In order to minimize the effect of variation associated with the size of the district, the 136 districts were subdivided into three subsamples of 69 districts with enrollments of 25,000 to 50,000, 43 districts of enrollments of 50,000 to 100,000, and 24 districts of greater than 100,000. Results for these samples are also shown in Table 1.[7] In

---

[4] Average monthly salaries by school districts are available from the *Census of Governments*, but the average salary for all classes of teachers would not be appropriate for the present study. An equation attempting to explain variation in average salary would have to include measures of the educational attainment of the faculty and the seniority of the faculty. By choosing the entry salary, we avoid both of these and also escape a possible problem of multicollinearity.

[5] Entry salaries, percent of revenues raised locally and effective tax rates are reported by the National Education Association in their *Selected Statistics*, and *Salary Schedules*. School districts are reported in the *Census of Governments, 1967*, and *Education Directory, 1967*. County per capita income figures are published by Standard Rate and Data Service. In preliminary work on this paper, turnover rates and the percentage of teachers holding masters degrees were included as explanatory variables. However, neither was significant.

[6] In addition to the equation in Table 1, several other formulations containing the monopsony variable and one or two of the others were regressed. With the exception of the large-district sample, the monopsony variable was highly significant in each of these equations.

[7] As an additional test for the effects of the three district sizes, we regressed the entire 136 observations for entry salary on the four variables listed above and on two dummy variables, D_1 and D_2, taking values of 1 and 0 according to whether the observation was or was not in a certain size group. D_1 refers to the intermediate sized districts, and D_2 to the large districts. The regression equation was:

$$Y = 4138.76 + \underset{(21.78)}{118.04\overset{a}{X_1}} + \underset{(64.70)}{230.19\overset{a}{X_2}} + \underset{(6.72)}{5.34X_3}$$

$$+ \underset{(2.02)}{4.63\overset{b}{X_4}} + \underset{(58.20)}{22.90D_1} + \underset{(71.11)}{109.97\overset{d}{D_2}}$$

$$R^2 = .49$$

where X_1, X_2, X_3, and X_4 refer, respectively, to the monopsony, per capita income, revenue from local

TABLE 1—ANALYSIS OF REGRESSION EQUATIONS–DEPENDENT VARIABLE IS BEGINNING TEACHER'S CONTRACTUAL SALARY

No.	Enrollment (1000's)	Constant	Logs of Districts	Per Capita Income	Percent Rev. Local	Tax Rate	Mid-West	South	West	R^2
1	Above 25 ($N=136$)	4132.00	116.04[a] (21.67)	242.32[a] (64.30)	4.90[b] (2.01)	4.30 (6.68)				.47
2	25–50 ($N=69$)	4174.42	163.34[a] (32.91)	211.87[b] (82.19)	5.86[b] (2.92)	−5.71 (8.69)				.51
3	50–100 ($N=43$)	4082.58	114.96[a] (34.93)	295.21[b] (138.89)	.42 (3.37)	19.42[d] (11.64)				.57
4	Above 100 ($N=24$)	4047.79	56.50 (51.64)	185.56 (178.60)	6.23 (5.14)	36.61[d] (25.17)				.39
5	Above 25 ($N=136$)	4563.86	74.64[a] (22.30)	190.50[a] (61.70)	4.96[b] (1.95)	−4.41 (6.51)	−111.50[d] (83.72)	−329.68[a] (85.19)	−16.60 (82.61)	.56
6	25–50 ($N=69$)	4714.28	110.52[a] (35.72)	174.42[b] (77.00)	2.84 (2.80)	−14.17[d] (8.68)	87.16 (133.65)	−354.46[b] (146.45)	25.72 (134.18)	.63
7	50–100 ($N=43$)	4414.28	94.03[b] (36.71)	220.41[d] (137.28)	1.64 (3.40)	18.94[c] (11.19)	−222.60[d] (138.58)	−263.93[b] (127.49)	−89.33 (124.61)	.64
8	Above 100 ($N=24$)	4497.87	1.24 (79.11)	15.26 (182.70)	13.96[b] (6.34)	7.02 (27.63)	−67.06 (251.62)	−236.49 (213.85)	441.09 (345.70)	.58

Notes: Figures in parentheses are standard errors.
a, b, c, and d represent significance from zero at the .005, .025, .05, and .10 levels, respectively. Data for 1966–67.

equations 2 and 3, the monopsony variable is again significant at the .005 level. In equation 2, the tax rate coefficient does not have the expected sign, and it is not significantly different from zero. For the districts with fewer than 100,000 students, the degree of competition in the labor market appears to be an important determinant of the entry level salary. Even though the monopsony variable is not significantly different from zero in equation 4, we show below that this is still consistent with the hypothesis being tested.

We believe that the simple model above provides strong empirical evidence that teachers' salaries vary inversely with the degree of competition in the relevant labor market. However, it may be possible that the equations discussed thus far contain spurious correlations between salary and monopsony that are caused by omitted but important variables. One possibility is that the income variable might not adequately express differences in living costs. If regional variation in the cost of living is coincidentally but significantly correlated with the number of school districts, then the spurious correlation mentioned here might result. To test for this possibility, we added three dummy variables to our previous set to adjust for purely regional differences among four geographic areas. The results of these additions are presented in equations 5 through 8 in Table 1. These equations repeat the pattern of the first half of the table. The monopsony variable is significant in the total sample and in the two smaller district samples, but not in the large district sample. The addition of the three regional dummies merely serves to point out that teachers' salaries are lower in the South than in the Northeast, other things

---

sources, and effective property tax variables as used in Table 1.

TABLE 2—ANALYSIS OF LARGE DISTRICTS

	Constant	Percent of District Students in SMSA	Per Capita Income	Percent Rev. Local	Tax Rate	Mid-West	South	West	R^2
1	4732.43	-581.58^d (368.21)	+184.71 (170.72)	+2.11 (5.86)	+29.66 (22.99)				.42
2	4849.35	-494.07^d (360.99)	+2.88 (72.85)	$+9.92^d$ (5.94)	+15.89 (24.19)	+26.42 (174.94)	−115.89 (217.28)	$+518.55^c$ (249.13)	.62

Note: Standard errors are in parenthesis.
a, b, c, and d represent significance from zero at the .005, .025, .05, and .10 levels, respectively.

being equal. The monopsony hypothesis is not weakened significantly by considering these variables.

As already noted, the monopsony variable is not significant in the sample consisting entirely of large school districts. One possible explanation for the lack of significance for the monopsony variable, we believe, can be found in the very size of these districts. Since all of them have over 100,000 students (the mean is 191,600), these districts should possess a considerable amount of monopsony power even if there are many other districts (which are likely to be much smaller) in the same county. Accordingly, we repeated the regression equation for these districts substituting a concentration variable for the log of the number of districts. In this case, the concentration variable is the percentage of total students in the Standard Metropolitan Statistical Area that is enrolled in the district in question. We anticipate that the entry salary is negatively correlated with the concentration ratio and therefore expect a negative coefficient for this variable. The results appear in Table 2. Equation 1 excludes and equation 2 includes the regional variables.

The same result as above holds. The entry salaries in the larger school districts are negatively correlated with labor market concentration. The coefficient for this term exceeds its standard error in both equations and is statistically significant in both at the .10 significance level.

III. Conclusions

From these results it is quite clear that competition in the local market for teachers does play a consistent and significant role in the determination of salary levels. The implications for policy and strategy here are quite interesting. First, several popular proposals for school district decentralization have recently been offered. The implementation of such plans would increase the number of bargaining units in the labor markets and, according to our findings, result in increased pressures for higher salaries. Teachers would stand to benefit at the cost of greater financial commitments on the part of the taxpayer.[8] On the other hand, any movement toward consolidation would increase the bargaining power of the school administrators and correspondingly weaken the position of the teachers. The natural recourse for teachers, in this case, is strong unionization which might be quite difficult in districts of such large size.

REFERENCES

C. S. Benson, *The Economics of Public Education,* 2d ed., Boston 1968.

D. Hellriegel, W. French, and R. Peterson, "Collective Negotiations and Teachers: A

---

[8] To say that taxpayers lose under decentralization overstates the case somewhat. Although an increase in the number of school districts may increase financial pressures for wage increases, smaller districts might be more effective in constructing optimally sized schools and reap offsetting economies through increased efficiency (or perhaps lose even more through decreased efficiency). However, a compelling argument for school decentralization is that it permits educational curricula to be more closely tailored to the differing individual preferences of individual communities. Such advantages may more than offset increased personnel costs of the type studied in this paper.

Behavioral Analysis," *Ind. Labor Relat. Rev.*, Apr. 1970, *23*, 380–96.

W. Z. Hirsch, "Demand and Supply of Teachers in California," *Calif. Manage. Rev.*, fall 1967, *10*, 27–34.

H. Kasper, "The Impact of Collective Bargaining on Public School Teachers," *Ind. Labor Relat. Rev.*, Oct. 1970, *23*, 57–72.

R. L. Johns and E. L. Morphet, *Financing Public Schools*, Englewood Cliffs 1965.

J. A. Kershaw and R. N. McKean, *Teacher Shortages and Salary Schedules*, New York 1962.

J. H. Landon, "The Effect of Product-Market Concentration on Wage Levels: An Intra-Industry Approach," *Ind. Labor Relat. Rev.*, Jan. 1970, *23*, 237–47.

J. Robinson, *Economics of Imperfect Competition*, London 1934, 211–28.

R. S. Thornton, "The Effects of Collective Bargaining on Relative Teacher's Salaries," presented before the annual conference of the Southern Economic Assoc., Nov. 1970.

National Education Association, *Salary Schedules for Teachers*, 1966–67, Washington 1968.

———, *Selected Statistics of Local School Systems*, 1966–67, Washington 1968.

Standard Rate and Data Service, *Newspaper Circulation Analysis, 1967–68*, Skokie, Ill. 1968.

U.S. Department of Commerce, *Statistical Abstract of the United States, 1969*, Washington 1970.

———, *Census of Governments, 1967*, Washington 1968.

U.S. Department of Health, Education, and Welfare, *Education Directory, 1967*, Washington 1968.

Shortsightedness and Nonmarginal Pareto Optimal Redistribution

By A. MITCHELL POLINSKY*

This note further elaborates the recent discussion of "Pareto Optimal Redistribution" by Harold M. Hochman and James D. Rodgers (1969), the subsequent comments by Richard A. Musgrave and Robert S. Goldfarb, and the reply by the original authors (1970). It demonstrates that conceptually there are two categories of Pareto optimal redistribution, only one of which was recognized by Hochman and Rodgers, and that this oversight can be explained in terms of their reliance on J. de V. Graaff's incorrect elucidation of Paul A. Samuelson's concept of the utility possibility frontier. Recognition of the second category of Pareto optimal redistribution requires further differentiation between "primary" and "secondary" distribution suggested by Musgrave. Its introduction also gives rise to a new kind of market failure due to "shortsightedness."

I. The True " 'True' Welfare Frontier": A Correction of Graaff

The first step in the present argument is to observe the error in Graaff's diagrammatic formulation of the welfare frontier, first introduced to the English literature by Samuelson (p. 244): When the optimal conditions of production and exchange are satisfied, the frontier can be stated as "an implicit equation between the levels of well-being of the different individuals in the system." Samuelson wrote this function as:

$$(1)\qquad P(U^1, U^2, U^3, \dots, U^s) = 0$$

This says that we may specify at will the levels of well-being of all but one of the individuals, and the last one's well-being is uniquely determined. The essential shape of this possibility function depends of course upon the technological and other assumed restraints of the system as well as upon the tastes of the different individuals. [p. 244]

In a mathematical appendix, Graaff defines this function almost identically:[1]

> The welfare frontier, or utility-possibility function, is a relation between the individual utility indicators of the form . . . [of (1)] which tells us the maximum value any one can attain (in view of the technological restraint imposed by the social transformation function), given the values of the remaining ones. [p. 73]

While Gaaff's mathematical statement is correct, his graphical illustration and verbal argument supporting it are incorrect in two respects, one of which he parenthetically acknowledges. Graaff drew the welfare frontier (p. 60 Figure 9a) as in Figure 1, with (from left to right) an initial downward sloping section, followed by an upward sloping section, another downward sloping section, and finally, an upward sloping section.[2] Strictly interpreted, as Graaff acknowledges, it is incorrect to include upward sloping sections in the frontier:

> The importance of the direction of slope of the frontier is considerable, for (if we confine ourselves to Pareto W's) we can rule out those parts which slope upwards as positions of maximum welfare. When it slopes upwards we can make everybody better off. In a sense, therefore, *we can speak of the (possibly discontinuous) downward-sloping section as the*

---

* Graduate student in economics at the Massachusetts Institute of Technology. This paper was written while I was a member of the research staff of the Urban Institute. I wish to acknowledge support from the Woodrow Wilson National Fellowship Foundation, under an Independent Study Award, and the National Science Foundation, under Grant GS-2805 to the Urban Institute. Helpful suggestions were received from Harold M. Hochman, an anonymous referee, and the editor of this *Review*.

[1] He mathematically states the equation in an implicit form identical to Samuelson's formulation.

[2] For expositional purposes I have slightly altered his diagram. The analysis is unaffected. Ignore momentarily the significance of the points on the diagram.

> "*true*" *welfare frontier*; and we can also speak of a sense in which the distribution of real income is "not optimal" on upward-sloping sections of the frontier. [p. 62, emphasis added]

Graaff's true welfare frontier is illustrated in Figure 2, where the upward sloping sections *CE* and *HI* have been eliminated.

Graaff does not go far enough in ruling out sections of his original welfare frontier. There are certain *downward sloping* sections which should also be eliminated. Any point on segment *AC* is Pareto-inferior to points on another attainable downward sloping section of the frontier. Consider point *B* in Figure 2. All Pareto-superior points to *B* are indicated by the region in the northeast quadrant of the axes drawn through *B*. It is obvious that any point on the segment *EG* would leave both individuals better off (or, at *G*, would make one better off without making the other worse off). This is true for every point on the downward sloping segment *AC* (although at *A* the only possible change is to hold one individual's utility constant while increasing the other's). Hence the true "true" welfare frontier, illustrated by Figure 3, is composed of downward sloping segments, but not *all* downward sloping segments, as Graaff thought, since it is possible to make everyone better off by moving away from points on some such segments.

While the above analysis is technically correct, it may not be of much interest by itself. However, when interpreted in terms of the discussion of Pareto optimal redistribution, it has some interesting implications.

II. Two Kinds of Pareto Optimal Redistribution

In their original paper, Hochman and Rodgers interpret their analysis in terms of a welfare or utility possibility frontier:[3]

> An alternative way of viewing the problem posed in this paper is in terms of the utility possibility function, a construction frequently employed in welfare economics. The existence of external economies can result in this function having upward sloping portions, positions which cannot be efficient in the Pareto sense. *The problem we analyze is one of moving, by means of redistributive transfers, from such an inefficient point to a point where the function no longer slopes upward.* [p. 543–44, fn. 5, emphasis added]

This statement can be technically corrected in the same way that Graaff's diagram was amended. However, in so doing, a new variation of Pareto optimal redistribution is suggested. Furthermore, this kind may be a better description of some real world phenomena.

The phenomenon which Hochman and Rodgers discussed will be designated "marginal" Pareto optimal redistribution; it reflects benevolent interdependence among utility functions, *given a particular initial distribution of income* (between, say, a rich Mutt and a poor Jeff, the two mythical inhabitants of their two-person world). Given an initial distribution in which Mutt and Jeff start from a point like *D* on the boundary of the attainable welfare set (Figure 4),[4] Mutt will voluntarily make transfers to Jeff until point *E* on the boundary is reached.

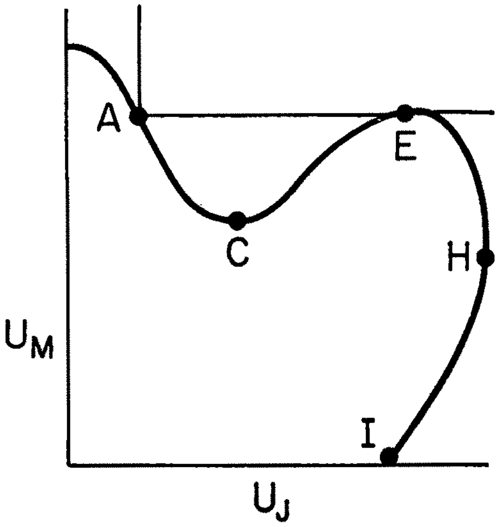
The second kind of Pareto optimal redistribution, which will be called "nonmarginal," corresponds to those downward sloping portions of the boundary of the frontier which are dominated by other downward sloping portions of the boundary. In Figures 1 and 2, the segment *AC* represents situations in which nonmarginal Pareto optimal redistribution is possible.

Before proceeding, it will be useful to provide an economic interpretation and

---

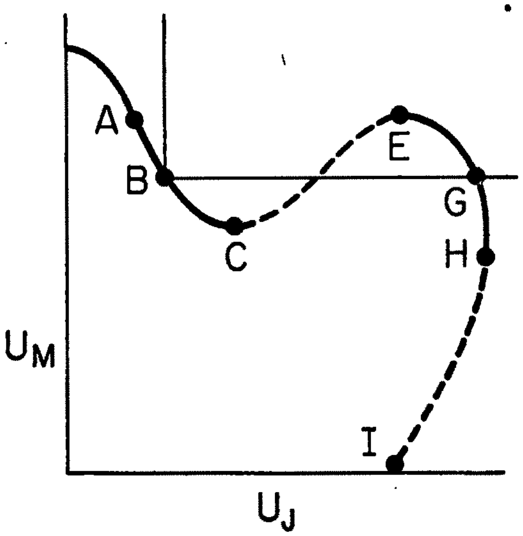
[3] In their later reply (1970), they do not use the terms "downward sloping" and "upward sloping" but rather (and correctly) "efficient" and "inefficient." However, they still identify, p. 999, Pareto optimal redistribution as involving "a movement from an inefficient (positive sloped) to an efficient portion of this locus [the welfare frontier]." That some points on downward sloping sections may be inefficient is still not recognized explicitly.

[4] Terminologically, it will be helpful to speak of an "attainable welfare set" (see Figure 4), describing all possible combinations of welfare given the technological and other constraints on the system. Then the proper welfare frontier consists of all points on the boundary of the attainable welfare set which are not dominated by (are not Pareto inferior to) other points on the boundary of this set. This definition of the welfare frontier will avoid Graaff's error.



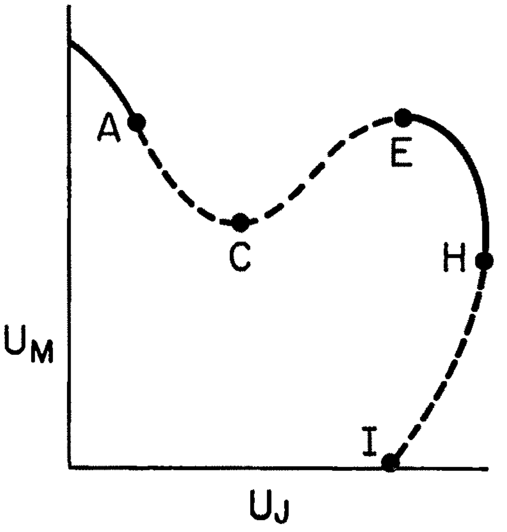
Graaff's Welfare Frontier

FIGURE 1



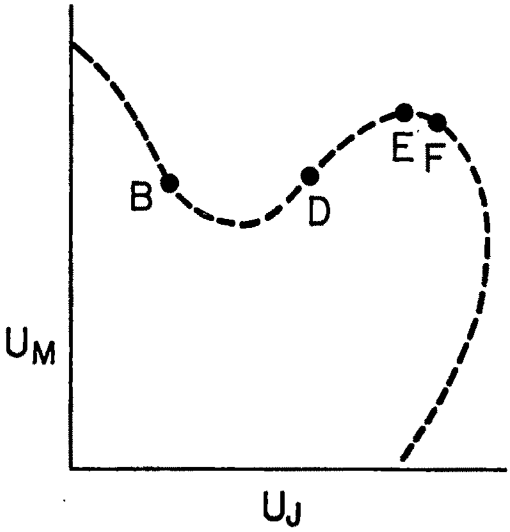
Graaff's 'True' Welfare Frontier

FIGURE 2



The True 'True' Welfare Frontier

FIGURE 3



The Attainable Welfare Set

FIGURE 4

justification of the boundary characteristics necessary and/or sufficient to account for both types of Pareto optimal redistribution. The "normal" case is still for the frontier to be monotonic downward sloping, thereby precluding either type of redistribution. Occasionally, however, the boundary may slope upwards. It will always be proper to equate these sections of the boundary with Hochman's and Rodgers' marginal redistribution. Upward sloping sections can be accounted for in three ways, set forth by Graaff:

> (1) The first is where the social transformation function [production possibility frontier] depends markedly on the

> distribution of wealth. If workers are underfed, a redistribution of wealth in their favor may so increase efficiency that everybody benefits; . . . (2) External effects in consumption can bring about the same result; . . . (3) In a larger community [three or more persons], moreover, external *dis*economies can give the frontier an upward slope. [p. 61]

In a mathematical appendix, Graaff derives these conditions formally, pp. 73–74. While Hochman and Rodgers framed their analysis in terms of the second explanation, they need not have been so restrictive.

While upward sloping sections of the boundary are both necessary and sufficient to account for marginal Pareto optimal redistribution, their existence is only necessary for the nonmarginal case, which requires downward sloping sections that are dominated by other downward sloping sections. It is evident that something more than a frontier which begins (left to right) with an upward sloping section, is followed by a downward sloping section, and ends with another upward section which meets the horizontal axis, is needed to explain nonmarginal redistribution. A sufficient condition, and one which is most attractive economically (see below), is for the frontier to begin with a downward sloping section, followed by an upward section, and so forth, as illustrated in Figures 1–4. This *is* in fact how Graaff drew his frontier, p. 60, and there does not appear to be anything in his mathematical conditions which precludes this.

It is necessary to add some substance to this still economically empty condition. Two verbal arguments will be presented here; they are developed more formally in the Appendix. The first argument is based on the transformation function explanation of upward sloping sections of the frontier. Starting at (or very near) one corner of the welfare frontier, if a small transfer is made (small being defined as not enough to assume subsistence), then the poorer person will not subsist and will still not be able to contribute to the productive activity of the society.[5] However, as soon as the transfer exceeds the subsistence level, both individuals can be made better off, up to the point where further transfers do not affect the production possibility frontier. Marginal transfers are wasted, while nonmarginal transfers are mutually beneficial, up to a point.

The second argument to account for nonmarginal redistribution is based on consumption externalities. Since it is more in the spirit of the Hochman and Rodgers paper, it is appropriate to frame it in a simple Mutt and Jeff example. Suppose the initial welfare positions correspond to a point such as B in Figure 4. If Mutt gives Jeff a dollar (a marginal change), Jeff merely "wastes" it (from Mutt's viewpoint), say, by consuming liquor. Mutt becomes worse off while Jeff is made better off. *At the margin*, redistribution is competitive (or "primary" in Musgrave's terminology). However, suppose that Mutt gives Jeff five thousand dollars (a nonmarginal change). Jeff now becomes an upstanding member of the community, pleasing Mutt greatly, so that their post-transfer utilities place them at a point such as F in Figure 4. Both are made better off. *Beyond the margin*, redistribution is benevolently interdependent, i.e., Pareto optimal.

Either argument is sufficient to account for a downward sloping section which is dominated by another downward sloping section. Furthermore, it can be argued (although not conclusively) that nonmarginal Pareto optimal redistribution is a more realistic description of real world behavior than marginal Pareto optimal redistribution. Many individuals would not be willing to give a beggar a dime, yet do voluntarily give hundreds of dollars to a large organized charity. Since the charity's resources usually are not spread so thinly that each ultimate beneficiary receives a dime, contribution of a large sum may reflect in part the hypothesized nonmarginal Pareto optimal redistribution. One may still justifiably question this and ask why the individual was not initially willing to make a nonmarginal transfer to the beggar. This can be explained as follows: The transfer he could make unilaterally to a single individual would be less than the amount needed to "jump the gap" to a Pareto optimal point.[6] Therefore, only through a cooperative effort, such as an organized charity or an institutionalized tax structure, could society reach a Pareto "optimal point.[7]

---

[5] The argument need not be formulated in such extreme "life and death" terms. It is more reasonable that a person would subsist, but would contribute so little to the society's productive activities that the transfer would still be competitive, thereby accounting for an initial downward sloping section.

In practice, it is nearly impossible to distinguish between the different kinds of redistribution. On this point all parties agree.[8] One can therefore only offer suggestive arguments for the significance of nonmarginal Pareto optimal redistribution, just as in the case of the more conventional marginal redistribution argument of Hochman and Rodgers. Conceptually, at least, nonmarginal Pareto optimal redistribution is on as firm a foundation as the marginal form.

The present development therefore necessitates further differentiation between primary and secondary redistribution suggested by Musgrave. He defines primary redistribution as "not a matter of voluntary giving, but of taking" (p. 991) while secondary redistribution is a matter of voluntary giving.[9] Nonmarginal Pareto optimal redistribution has characteristics which reflect both of Musgrave's criteria. It is primary at the margin, but secondary beyond the margin. However, since it would involve a voluntary transfer if all involved were aware of the shape of the attainable welfare set, it properly comes under Musgrave's secondary category. It is now necessary to speak of three kinds of redistribution: primary, marginal secondary, and nonmarginal secondary.

While the nonmarginal secondary redistribution would be voluntary in a world without costs of information or bargaining, it is quite possible that, realistically, the parties involved will not see beyond the margin. The implications of this problem will now be explored.

III. Shortsightedness As a Market Failure

In a two-person world, Hochman and Rodgers found no problem with the attainment of a Pareto optimum. If there is benevolent interdependence, Mutt will voluntarily make a transfer to Jeff, making both better off.[10] In an N-person world, Pareto optimal transfers may or may not come about, depending on one's assumptions. If, as Hochman and Rodgers suggest, one assumes that

> . . . each benefactor's benefits derive only from transfers that he makes himself and not from transfers which other persons make, then simple aggregation of pairwise transfers is the proper method of analysis in the N-person case and [their] straightforward extension of the two-person model to the case of a community-at-large is appropriate. [1970, p. 1000]

On the other hand, if the transfer from a particular Mutt to a particular Jeff yields spillover benefits to other Mutts, then Hochman and Rodgers agree that "When there is a large number of potential benefactors, the attainment of a Pareto optimum through voluntary action is difficult, inasmuch as each participant has a private incentive to remain a 'free rider' . . ." (1970, p. 1000).

Goldfarb's comment, which proposes a slightly altered specification of the Hochman and Rodgers argument, suggests another impediment to the attainment of a Pareto optimum through private behavior in an N-person world. Pareto optimal transfers may not come about voluntarily if (a) each Mutt's marginal utility from the last dollar of own-consumption is greater than his marginal utility from giving that dollar to a Jeff; (b) each Mutt's marginal utility of giving a dollar to a Jeff is greater than zero; and (c) there are a "sufficient" number of Mutts.[11] Under these conditions, no one would voluntarily make transfers, although all would be better off if *all* rich persons, acting collectively, imposed a redistribution scheme on themselves.

---

[6] By definition, this phenomenon cannot be illustrated in the two-person diagrams of Figures 1–4.

[7] In addition to the "market failures" discussed in Section III below, it is appropriate to acknowledge here some further problems which will henceforth be assumed away. If jumping the gap requires a large number of contributors, and if different rich individuals have different "philanthropic preferences," just as they have different preferences for other goods, then optimal transfers may not result because of the cost of the charity itself. In such cases, the information costs of finding out who else would be willing to contribute to which charity, and the transaction costs of collecting and disbursing the contributions may be quite large relative to the actual value of the transfers.

[8] See Hochman and Rodgers (1970, p. 999) and Musgrave, p. 992.

[9] Under certain conditions, to be discussed in more detail in Section III, both nonmarginal and marginal secondary redistribution may not come about voluntarily because of market imperfections. With respect to the latter category this was recognized by Musgrave, p. 992, and Goldfarb, p. 994.

[10] Hochman and Rodgers (1969, p. 543). A similar argument is made by Graaff, p. 61.

With the introduction of nonmarginal Pareto optimal redistribution, another kind of market failure can occur, even in a two-person world. If individuals behave at the margin, i.e., if they are shortsighted as is commonly assumed, they may not perceive Pareto optimal points beyond the margin. In the N-person case, the problem is further complicated because the free-rider phenomenon compounds the problem posed by shortsightedness. Movements from points like B in Figure 4 are unlikely to occur on a voluntary basis. If the Mutts of the world are completely shortsighted (look no further than the margin), no transfer will be made voluntarily. Note that the free-rider phenomenon is not then perceived as an issue by the Mutts since each does not gain by a *marginal* transfer. However, to the extent that the Mutts realize that Pareto optimal points exist somewhere beyond the margin, the free-rider problem becomes operative, thereby still resulting in no move from B voluntarily.

The implication of this phenomenon for public policy is even stronger than under the case of marginal secondary redistribution. Previously, there was no justification for public intervention, on efficiency grounds, in a two-person world; such is not the case now if individuals are shortsighted. It now becomes desirable for the individuals to voluntarily form a government which would "coerce" them into behaving Pareto optimally. In the N-person case, the shortsightedness argument is still valid, but is compounded by the free-rider aspects of jumping the gap to a Pareto superior point. If the gap can be jumped by transfers from some subset of rich individuals, and if benefits accrue to everyone, then there will be an incentive to let the others pay. If a Pareto optimal point is to be attained, public intervention will very likely be necessary.

IV. Conclusion

In a world where there is a benevolent interdependence among individual utility functions, transfers can be made which are Pareto optimal. Hochman and Rodgers implicitly formulated this problem so that individual behavior at the margin will lead to a Pareto optimal solution (abstracting from the free-rider problem). It is now apparent that the phenomenon of Pareto optimal redistribution is somewhat more complicated.

There may exist regions on the boundary of the welfare set from which small movements are not Pareto optimal, but large movements are. Moreover, the shortsighted behavior of individuals acting at the margin will not lead to an attainable Pareto optimal point, both in the two-person and N-person cases, even abstracting from the general N-person problem of free-riders and the special case suggested by Goldfarb. As a result, and to the extent nonmarginal secondary redistribution is significant, the redistribution of income through public institutions can be justified on efficiency grounds even more strongly than was argued by Hochman and Rodgers.

APPENDIX

It was seen in Section II that the existence of marginal Pareto optimal redistribution did not imply the nonmarginal case. The two verbal arguments used to account for the nonmarginal variant will now be developed graphically. A third explanation, based on external *dis*economies in consumption, re-

---

[11] There must be at least two rich persons, as well as at least one poor person. The Goldfarb phenomenon is therefore not possible in a two-person world.

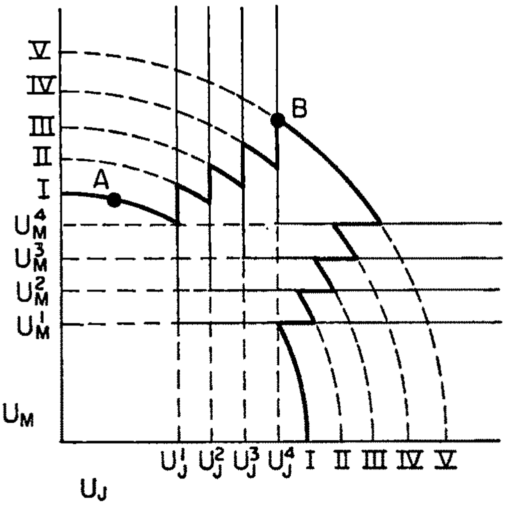


FIGURE 5

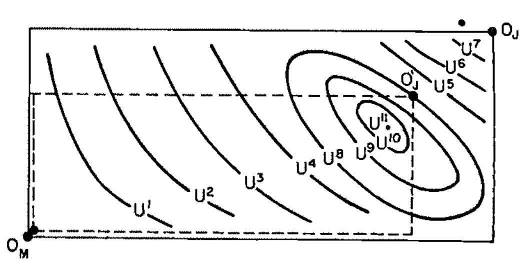


FIGURE 6

quires at least a three-person world and will not be considered here.

In the first case, consider a symmetrical two-person world (for graphical convenience) in which the productive capacity of each individual increases discretely with increases in his utility. Assume that below some level of utility he does not produce at all (but that he continues to survive), and that beyond some level of utility further consumption does not affect his productivity. The implications of this analysis are illustrated in Figure 5. Below the level of utility U_J^1, Jeff (for example) does not produce. Any output must result from Mutt's productivity, which corresponds to the possibility locus *I-I*. For example, at point *A* only Mutt is producing something, although an *involuntary* transfer has been made to Jeff (say through a tax system). However, if Mutt should continue to give Jeff more of his product, then eventually (beyond U_J^1) Jeff will also be capable of producing more goods. The combined capabilities of Mutt and Jeff, when Jeff's utility is in the range U_J^1 to U_J^2, is represented by the locus *II-II*. The remaining sections of the boundary are constructed similarly, up to the point where the production and utility possibility frontiers are not affected by the utility of the producers (i.e., up to U_J^4 and U_M^4, respectively). The discrete case was used for clarity. Realistically, the transitions would be smoother, and the boundary would consequently smooth out. The downward sloping sections at either corner correspond to nonmarginal redistribution since movements from points like *A* to points like *B* (on another downward sloping section) leave both individuals better off.

The second case is illustrated in Figure 6, which is an extension of a graphical device used by Kenneth Boulding (Figure 3, p. 70).[12] Boulding employed the Edgeworth box diagram to illustrate indifference maps which reflect interdependence between utility functions such that voluntary transfers would be made by one individual starting from either origin of the box. His diagram actually is drawn similar to what is included in the *dotted* box of Figure 6. If one started at the dotted origin O_J', in which Mutt had control of all of the society's resources, he would voluntarily transfer goods to Jeff since by so doing he could increase his own welfare from U^8 to U^{11}. While Jeff's indifference curves are not illustrated, we can assume he would also be made better off. Thus, starting at one corner of the utility possibility frontier, voluntary *marginal* redistribution would occur. Now consider a slightly altered case illustrated by the entire *solid* box of Figure 6. It is straightforward to observe that if Mutt now began with all of both goods, at origin O_J, he would not make voluntary transfers to Jeff if he did not see beyond the margin. Small transfers would make him worse off and Jeff better off. However, a sufficiently large transfer could increase

---

[12] I have only included those aspects of Boulding's diagram which are necessary to illustrate the present argument. The indifference curves are drawn for Mutt only, and increase such that $U^1 < U^2 < \dots < U^{11}$.

Mutt's utility to U^{11}. The corresponding possibility frontier would have (starting at Mutt's corner) an initial downward sloping section (from U^7 to U^5), an upward sloping section (U^8 to U^{11}, moving right to left), and a downward section (U^{11} to U^8, continuing to the left) which dominates the original downward section. Hence, this illustrates the possibility of nonmarginal redistribution.[13]

REFERENCES

K. E. Boulding, "Notes on a Theory of Philanthropy," in F. G. Dickinson, ed., *Philanthropy and Public Theory*, New York 1962, 57–71.

R. S. Goldfarb, "Pareto Optimal Redistribution: Comment," *Amer. Econ. Rev.*, Dec. 1970, *60*, 994–96.

J. de V. Graaff, *Theoretical Welfare Economics*, Cambridge 1957.

H. M. Hochman and J. D. Rodgers, "Pareto Optimal Redistribution," *Amer. Econ. Rev.*, Sept. 1969, *59*, 542–57.

——— and ———, "Pareto Optimal Redistribution: Reply," *Amer. Econ. Rev.*, Dec. 1970, *60*, 997–1002.

R. A. Musgrave, "Pareto Optimal Redistribution: Comment," *Amer. Econ. Rev.*, Dec. 1970, *60*, 991–93.

P. A. Samuelson, *Foundations of Economic Analysis*, New York 1947.

---

[13] For more detail on the abnormal characteristics of indifference maps in an Edgeworth box framework when utility functions are interdependent, the reader is referred to the interesting article by Boulding. Further discussion here is beyond the scope of this note.

A Set of Explicit Homothetic Production Functions

By BENJAMIN WOLKOWITZ*

Representing the elasticity of substitution, returns to scale, and technological change in a more general way than can be accomplished with either a Cobb-Douglas or *CES* production function has been the subject of several recent articles. In particular, deriving production functions indirectly by specifying the relationship between key variables of the production process has resulted in several new forms that are more general than either of the two preceding forms, (see Ryuzo Sato and Martin Beckmann, and Beckmann and Sato). Also, the property of homotheticity has been used to derive a numerical measure of variable returns to scale (see Simone Clemhout, and Arnold Zellner and Nagesh Revankar). This paper presents a set of explicit production functions incorporating homotheticity so as to represent a production process with variable returns to scale while allowing for varied behavior of the elasticity of substitution. The Cobb-Douglas and *CES* production functions, which are characterized by a constant elasticity of substitution and uniform returns to scale, are included as special cases of this general set of production functions.

I. The Basic Relationship

The implication of homotheticity when applied to a two-factor production function is that the slope of an isoquant, $\Psi(K/L)$, is a unique function of the slope of the ray which intersects it, i.e., the capital-labor ratio. Furthermore, the slopes of all isoquants are identical along a given ray. (See Figure 1.)

This relation can be stated mathematically as

$$MRTS = \Psi\left(\frac{K}{L}\right) \tag{1}$$

where

$MRTS$ = marginal rate of technical substitution
K = capital
L = labor
Ψ = general functional notation

By specifying $\Psi(K/L)$, isoquants can be generated which will have particular curvature properties, and the explicit production function can be derived by solving the differential equation implied by equation (1). The solution to equation (1) must be homogeneous of degree one in its arguments; however, it is quite conceivable that production functions homogeneous of some other degree will have isoquants of identical curvature.[1] Therefore, a transformation of the linearly homogeneous form must be considered before the final production function can definitely be identified. Generally, the solution to equation (1) is

$$Y = F[g(K, L)] \tag{2}$$

where

Y = output
g = linearly homogeneous function
F = monotonic transformation

which is the form of the class of homothetic production functions.

II. Specification of Ψ(K/L)

In order to specify a final homothetic production function, it is necessary first to solve for the linearly homogeneous form, Z, and then to specify the F transformation. To

---

* Assistant professor of economics, Tulane University. I would like to thank Ryuzo Sato, John R. Moroney, Tetsunori Koizumi, Mark Daniels, Mark Schupack, and an unknown referee for their suggestions and criticisms. Financial support was supplied by a Woodrow Wilson Doctoral Dissertation Fellowship. Responsibility for the final results remain with me.

[1] Since equation (1) is a linear differential equation of the first order, it can be solved by finding an integrating factor, μ, which will convert it to an exact differential equation. However, μ is not unique, and $\mu F(g)$ will also be a proper integrating factor yielding a solution to equation (1) as long as $F(g)$ is nonvanishing. (See E. L. Ince, pp. 27–28.)

derive Z, the form of $\Psi(K/L)$ is substituted into[2]

$$Z = Le^{\int \phi(x)dx} \tag{3}$$

where $x = K/L$

$$\phi(x) = \frac{1}{x + \Psi(x)}$$

There are two conditions that govern the choice of reasonable $\Psi(K/L)$ forms. First, $\Psi(K/L)$ must be monotonically increasing in order for the resulting production functions to satisfy the law of diminishing marginal rate of technical substitution.[3] Secondly, $\Psi(K/L)$ must yield a closed solution when substituted into equation (3). There are six forms that satisfy both conditions, and they allow for varied behavior of the elasticity of substitution as factor proportions are altered. Following are the forms and a discussion of the intermediate production functions.

A. Linear Form: Cobb-Douglas Type

$$\Psi(x) = a + bx \tag{4}$$

Substituting into the integral form for Z yields

$$Z = Le^{\int \frac{dx}{a+bx+x}} \tag{5}$$

Integrating

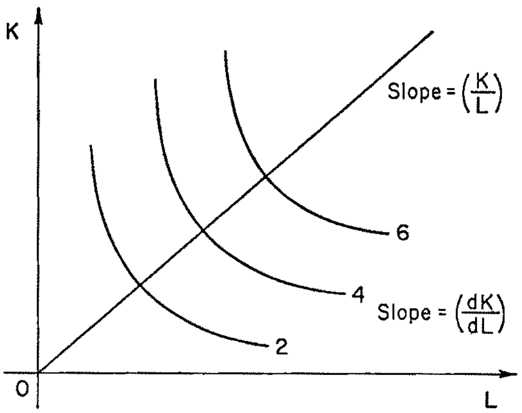

K
Slope = $\left(\frac{K}{L}\right)$
6
4
Slope = $\left(\frac{dK}{dL}\right)$
2
0
L

FIGURE 1

$$Z = Le^{1/(1+b)\ ln[(1+b)x+a]} \tag{6}$$

which can be written as

$$Z = L[(1 + b)x + a]^{1/1+b} \tag{7}$$

Rewriting this expression in terms of capital and labor yields the intermediate production function

$$Z = [(1 + b)KL^{b} + aL^{1+b}]^{1/(1+b)} \tag{8}$$

where b must be greater than zero and a can be of any sign as long as

$$(1 + b)KL^{b} > aL^{1+b}$$

The elasticity of substitution of this form is functionally related to the capital-labor ratio when $a \neq 0$. For the case when $a > 0$, the elasticity of substitution is greater than one, and inversely related to the capital-labor ratio. When $a < 0$, the elasticity of substitution is less than one, and directly related to the capital-labor ratio. However, if a is zero, a Cobb-Douglas production function results,[4] i.e., if $a = 0$

$$\sigma(x) = \frac{\Psi(x)}{x\Psi'(x)} = \frac{bx}{b \cdot x} = 1$$

---

[2] A similar derivation has appeared in Clemhout and Sato and Beckmann but is repeated here for convenience. Substituting in the definitions for the wage rate and the return to capital in equation (1) yields

$$\frac{g(x) - xg'(x)}{g'(x)} = \Psi(x) = -\left(\frac{\partial Z}{\partial L} \bigg/ \frac{\partial Z}{\partial K}\right)$$

which can be rewritten as $g(x)/g'(x) - x = \Psi(x)$. Then let $\phi(x) = 1/(x + \Psi(x))$. From $Z/L = g(x)$, we have

$$\frac{d}{dx} log\ g(x) = \frac{d}{dx} log\ \left(\frac{Z}{L}\right) = \frac{1}{x + \Psi(x)}$$

which has a solution $Z = Le^{\int \phi(x)dx}$

[3] $$\Psi'\left(\frac{K}{L}\right) = \frac{d(g_L/g_K)}{d(K/L)} > 0$$

defines a diminishing marginal rate of technical substitution which is equivalent to diminishing marginal productivity for linearly homogeneous functions.

[4] One definition of the elasticity of substitution is

$$\sigma(x) = \frac{d\left(\frac{K}{L}\right) \bigg/ \left(\frac{K}{L}\right)}{d\left(\frac{dK}{dL}\right) \bigg/ \left(\frac{dK}{dL}\right)}$$

Substituting equation (1) into the definition we have $\sigma(x) = \Psi(x)/x\Psi'(x)$. See Clemhout.

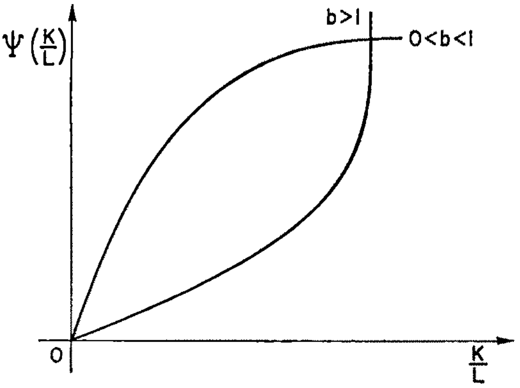


FIGURE 2

Thus the linear relationship results in an intermediate production function which is a Cobb-Douglas plus a linear term in labor when $a \neq 0$.

B. Log-Linear Form: *CES* Type

The initial relationship

$$\log \Psi(x) = a + b \log x \tag{9}$$

results in the intermediate production function

$$Z = L\left(\frac{x}{(1 + \alpha x^{b-1})^{1/(b-1)}}\right) \tag{10}$$

where $\log \alpha = a$ and b must be equal to or greater than 0.

The slope of the curve which plots the relationship between $\Psi(x)$ and x is dependent on the value of b. This may be seen in Figure 2. Note that $1/b$ is the value of the elasticity of substitution of the Z function. Therefore, the log-linear relationship results in a Z function which is a *CES* type.

C. Two Forms: Squared Independent Variables

The initial relationship may be either

$$\Psi(x) = a + Cx^2 \tag{11}$$

or

$$\Psi(x) = a + bx + Cx^2 \tag{12}$$

Both forms yield the intermediate production function

$$Z = \left(\frac{KL^{1/\alpha} - gL^{(1+\alpha)/\alpha}}{K - hL}\right)^{\alpha} \tag{13}$$

where

$$\alpha = \frac{1}{C(h - g)}; \qquad h - g \neq 0$$

h and g are the two roots of the quadratic equation which must be positive. Also, the larger of the two roots must be less than the smallest theoretically permissible capital-labor ratio.[5]

There are no restrictions on the coefficients of the estimated relationship but rather on the roots of the resultant quadratic equation.

The elasticity of substitution of the production function resulting from an initial squared relationship is inversely related to the capital-labor ratio when a and b are positive (C must be positive.) In this sense, the squared form is compatible with the linear form when a is greater than zero. However, if a and/or b are negative then the behavior of the elasticity of substitution cannot be determined until the magnitudes of a, b, and C are specified.

D. Semi-Log Form

The initial relationship

$$\Psi(x) = a + \beta \log x \tag{14}$$

results in the intermediate production function

$$Z = \left[aL^{1+\beta} + \beta L^{1+\beta} \log \frac{K}{L} + KL^{\beta}\right]^{1/1+\beta} \tag{15}$$

When $\Psi(x) = 0$, $\log x = -a/\beta$. A zero or positive point of intersection on the x-axis is expected, and this governs the expected hypothesized values in empirical work.

The elasticity of substitution of this intermediate production function is directly related to the capital-labor ratio so that when the capital-labor ratio rises, the elasticity of substitution rises.

---

[5] This restriction is pertinent to empirical analysis. If the smallest capital-labor ratio is less than the larger of the two roots, the intermediate production function will be negative.

E. Reciprocal Form

$$\Psi(x) = a + b/x \tag{16}$$

In order for the relationship to be monotonically increasing, the b coefficient must be negative.

When $\Psi(x) = 0$, the intersection on the x-axis is at $-b/a$; therefore, a must be greater than zero.

The intermediate production function, Z, is rather complex even though it is derived from a plausible relationship.

$$\log Z = \tfrac{1}{2} \log Q - R \log S \tag{17}$$

where

$$Q = x^2 + ax + (-b)$$

$$R = \frac{a}{2\sqrt{-b + \frac{a^2}{4}}}$$

$$S = \frac{x + \frac{a}{2} - \sqrt{-b + \frac{a^2}{4}}}{x + \frac{a}{2} + \sqrt{-b + \frac{a^2}{4}}}$$

The elasticity of substitution is a variable, but its behavior cannot be determined until the magnitudes of a and b have been specified.

Technical change parameters were deleted in the above derivations; however, their inclusion in empirical analysis would allow for a test to determine if technical change has been neutral or biased.

III. Specification of the F Transformation

Once the $\Psi(K/L)$ function is specified and Z is explicitly solved, it is necessary to specify the F transformation. The only required condition on this transformation is that it be monotonically increasing,[6] which indicates that the list of acceptable F transformations can be extended well beyond the list of acceptable $\Psi(K/L)$ functions.

All homothetic transformations result in production functions where the returns to scale are functionally related to the output level (provided that $F \neq 1$).[7] However, the final forms differ from one another in their sensitivity of returns to scale to changes in the output level. For example, the quadratic form is more sensitive than the linear form, but less sensitive than the cubic form. Determining the correct specification of returns to scale must necessarily be resolved through empirical analysis.[8]

REFERENCES

M. J. Beckmann and R. Sato, "On a Class of Production Functions Defined by Log-Linear Relationships," unpublished manuscript 1969.

S. Clemhout, "The Class of Homothetic Isoquant Production Functions," *Rev. Econ. Stud.*, Jan. 1968, *35*, 91–104.

E. L. Ince, *Ordinary Differential Equations*, New York 1965.

R. Sato, *The Theory of Economic Growth*, Tokyo 1968, (English translation forthcoming).

——— **and M. J. Beckmann**, "Neutral Inventions and Production Functions," *Rev. Econ. Stud.*, Jan. 1968, *35*, 57–66.

R. W. Shephard, *Cost and Production Functions*, Princeton 1953.

B. Wolkowitz, "Homothetic Production Functions: A Theoretical and Empirical Analysis," unpublished doctoral dissertation, Brown Univ. 1970.

A. Zellner and N. S. Revankar, "Generalized Production Functions," *Rev. Econ. Stud.*, Apr. 1969, *36*, 241–50.

---

[6] The reason for the monotonicity condition placed on F is clear from a consideration of the marginal products of the homothetic production function (see Zellner and Revankar).

[7] The magnitude of returns to scale of the homothetic production function is easily computed from *returns to scale* $= (\partial Y/Y)(Z/\partial Z)$.

[8] For comparative purposes when estimating the homothetic production function, an F transformation that yields a homogeneous production function, which by definition is homothetic, should also be included. That is

$$\log Y = C + \lambda \log Z + \lambda t$$

where Y is homogeneous of degree λ with respect to Z, and λ is equal to the returns to scale. This is the form used by Simone Clemhout in her empirical work.

NOTES

Editor's Announcement

Beginning January 1, 1972, the *American Economic Review* will require a submission fee of $10 to accompany each manuscript when it is first sent in for publication, regardless of size. However, manuscripts should not exceed fifty typewritten pages of text. Checks or money orders should be made payable to the American Economic Association.

The submission fee was approved by the Executive Committee of the AEA at its March 1971 business meeting. The fee has two purposes: First, it will help defray costs of handling and screening the submitted papers. Second, it will encourage authors to submit their work in more finished form and hopefully cut down on rejections which are based on the unfinished or preliminary quality of the papers.

Nominating Committee of The American Economic Association

In accordance with Section IV, paragraph 2, of the bylaws of the American Economic Association as amended last year, President-elect John Kenneth Galbraith has appointed a Nominating Committee for 1972 consisting of Charles L. Schultze, Chairman, Robert Eisner, William J. Fellner, Barbara Bergmann, Thaddeus Spratlen, and Robert Lekachman. Attention of members is called to the part of the bylaw reading, "In addition to appointees chosen by the President-elect, the Committee shall include any other member of the Association nominated by petition including signatures and addresses of not less than 2 percent of the members of the Association, delivered to the Secretary before December 1. No member of the Association may validly petition for more than one nominee for the Committee. The names of the Committee shall be announced to the membership immediately following its appointment and the membership invited to suggest nominees for the various offices to the Committee."

The Southeastern Regional Science Association is now the Southern Regional Science Association. Officers for 1971–72 are: president, Robert T. Miki, U.S. Department of Commerce; president-elect, William A. Schaffer, Georgia Institute of Technology; secretary-treasurer, Donald C. Darnton, Virginia Polytechnic Institute and State University.

The Association's annual meetings will be held in Williamsburg, Virginia, on April 13–14, 1972. Persons interested in participating are invited to contact the program chairman: Professor William A. Schaffer, School of Industrial Management, Georgia Institute of Technology, Atlanta, Georgia 30332.

The twenty-third annual meeting of the New York State Economics Association (NYSEA) will be held on April 21–22, 1972 at the State University of New York, College at Geneseo, New York. The guest speaker will be Kenneth Boulding.

Officers of NYSEA for 1971–72 are: president, David A. Marin, State University of New York at Geneseo; vice president, William N. Leonard, Hofstra University; secretary-treasurer, William B. Bennett, State University of New York at Buffalo; editor, Daniel Feinberg, New York City Community College.

The National Center for Health Services Research and Development is seeking research grant proposals from economists interested in working in the areas of health economics. Inquiries should be sent to Chief, Economic Analysis Branch, National Center for Health Services Research and Development, 5600 Fishers Lane, Parklawn Building, Rockville, Maryland 20852.

A Clearinghouse for Research on Civil Rights

The Center for National Policy Review is devoted to assuring that federal laws and policies in urban areas are implemented to protect the interests of minority groups and the poor. It is composed of lawyers and social scientists engaged in research and legal action on behalf of the principal civil rights groups in the country, including the NAACP, the Urban League, and the National Urban Coalition.

The Center will serve as a clearinghouse for social science research which has implications for federal policies on equality of opportunity. We solicit brief summaries of such research which is either completed or in progress. The principal concerns are with housing, credit practices, federal site selection, employment, and metropolitan school desegregation, as well as other areas such as welfare program administration, political participation, and revenue sharing. These research reviews will appear in a quarterly newsletter. The Center for National Policy Review is located at the School of Law, Catholic University of America, Washington, D.C. 20017, (202) 832-8525. The Director of Research is Samuel Gubins.

The National Science Foundation has funded the Laboratory for Political Research at the University of Iowa to publish a quarterly newsletter of information on the acquisitions of social science data archives in the United States. The newsletter will contain brief descriptions of new data sets recently acquired by archives and will also include names and addresses of appropriate persons to contact for those interested in acquiring specific data sets.

Individuals and institutions may be placed on the Newsletter mailing list by writing to John G. Kolp, Managing Editor, Newsletter of Social Science Archival Acquisitions, 321A Schaeffer Hall, University of Iowa, Iowa City, Iowa 52240.

A visiting lecturer program in statistics has been organized for the ninth successive year. The program is sponsored jointly by the principal statistical organizations in the United States, the American Statistical Association, the Biometric Society, and the Institute of Mathematical Studies. The National Science Foundation provides partial financial support. Leading teachers and research workers in statistics from universities, industry, and government, have agreed to participate as lecturers. Lecture topics include subjects in experimental and theoretical statistics, as well as in such related areas as probability theory, information theory, and stochastic models in the physical, biological, and social sciences.

The purpose of the program is to provide information to students and college faculty about the nature and scope of modern statistics, and to provide advice about careers, graduate study, and college curricula in statistics. Inquiries should be addressed to: Visiting Lecturer Program in Statistics, Department of Statistics, Southern Methodist University, Dallas, Texas 75222.

The Joint Committee on Eastern Europe of the American Council of Learned Societies and the Social Science Research Council wishes to announce the following programs:

Grants for post-doctoral research are offered to mature scholars for research in the humanities or social sciences relating to Albania, Bulgaria, Czechoslovakia, Hungary, Poland, Romania, Yugoslavia, East Germany since 1945, and post-Byzantine Greece. Small grants are awarded for research-related travel, research assistance, and maintenance for short periods of time. Grants to a maximum of $8,500 are also awarded for more extended periods. Deadline for receipt of applications: December 31, 1971.

Grants in support of conferences are made to meet part of the costs of conferences held in the United States for the advancement of research in the East European field, exclusive of Russian/Soviet studies. Normally a year's lead time for preparation of papers is considered desirable. Applications should be submitted by February 15, 1972.

For details of eligibility and information which *must* be supplied in requesting application forms, request an announcement brochure from: Office of Fellowships and Grants, American Council of Learned Societies, 345 East 46 Street, New York, New York 10017.

Papers are solicited for the fourth Amos Tuck School Seminar on Problems of Regulation and Public Utilities, sponsored by the American Telephone and Telegraph Company, which will be held in late August 1972, at Dartmouth College. The seminar consists of seven or eight sessions over a four-day period, and it will be attended by thirty conferees from academic institutions throughout the country. The sessions, each of which will focus on a single paper, provide a forum for the presentation and discussion of new ideas and innovative work in analyzing regulated firms, regulatory processes, and problems of public welfare.

We invite those with an interest in this area to suggest names of possible speakers or to apply themselves. The aim of the Seminar is to introduce the research both of young scholars beginning their careers and of scholars who have only recently turned their attention to public utility problems. Therefore, suggestions must be limited to scholars in these categories. Those chosen to present papers will receive up to $1,200 in support of research and writing, in addition to travel and living expenses during the Seminar. Please send suggestions and/or biographical information and samples of work, including a finished paper and a proposal to be developed for presentation, to the directors of the Seminar before February 29, 1972.

This is not a request for conferee applications. Application forms and further information will be mailed to institutions in March, 1972. The directors of the Seminar are: Professors Richard S. Bower and Willard T. Carleton, The Amos Tuck School of Business Administration, Dartmouth College, Hanover, New Hampshire 03755.

At the annual meeting of the American Association for the Advancement of Science, a session will be devoted to Contributions of Social Science Research in Education. The session will be held December 29, 1971 at the Holiday Inn, Penn Center, Philadelphia. The following papers will be read: "The Social Sciences and Their Contributions to Basic Research in Education," Henry David; "Contributions of Economics to Basic Research in Education," Finis Welch; "Contributions of Political Science to Basic Research in Education," Robert Lane; "Contributions of Anthropology to Basic Research in Education," John Singleton; and "Support for Social Science Research," Howard Hjelm.

Retirements

E. Kingman Eberhart, department of economics, College of Wooster.

Visiting Foreign Scholars

Michael Kidron, Oxford University: visiting professor of economics, University of California, Davis, Jan.–July 1972.

David H. Metcalf, London School of Economics: visiting associate professor, department of economics, Washington University, fall 1971.

Don Patinkin, Hebrew University: visiting scholar in banking and finance, University of Georgia, fall quarter 1971.

Promotions

William M. Baird: professor, department of economics, College of Wooster.

Theodore C. Bergstrom: associate professor, department of economics, Washington University, fall 1971.

Dale A. Berry: professor, department of economics, Drake University.

Charles W. Bischoff: associate professor of economics, Yale University.

Donald J. Brown: assistant professor of economics, Yale University.

Andrzej Brzeski: professor of economics, University of California, Davis.

Les O. Carson: associate professor, departments of economics and business administration, Augustana College, Sept. 1971.

David H. Ciscel: assistant professor, department of economics, Drake University.

Dean S. Dutton: associate professor, department of economics, Brigham Young University.

Paul H. Eggers: professor, departments of business administration and economics, Augustana College, Sept. 1971.

David Eiteman: professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

Arthur Geoffrion: professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

Peter J. George: associate professor of economics, McMaster University.

Richard Goodman: associate professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

William D. Gunther: associate professor of economics, University of Alabama.

Harold Kassarjian: professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

Donald W. Katzner: professor, economics department, University of Waterloo.

Arthur J. Kirsch: associate professor of economics, California State College at Los Angeles.

Charles G. Leathers: associate professor of economics, University of Alabama.

Steven An-Yhi Lin: associate professor of economics, Southern Illinois University at Edwardsville.

Steven Lippman: associate professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

N. Harvey Lithwick: professor, Carleton University, July 1, 1971.

William McKelvey: associate professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

Vahid F. Nowshirvani: associate professor of economics, Yale University.

Efthimios Pournarakis: associate professor of economics, University of Akron.

Georg Rich: associate professor, Carleton University, July 1, 1971.

Ulaganathan Sankar: associate professor, department of economics, University of Wisconsin, Milwaukee.

Byron G. Spencer: associate professor of economics, McMaster University.

Elias H. Tuma: professor of economics, University of California, Davis.

Leon L. Wegge: professor of economics, University of California, Davis.

Administrative Appointments

Faraj Abdulahad: chairman, department of economics, Manhattan College, July 1, 1971.

Murray Altmann: assistant secretary, Office of the Secretary, Board of Governors of the Federal Reserve System, July 1971.

Roy Bahl, International Monetary Fund: associate professor of economics; associate director, Metropolitan Studies Program, Syracuse University.

Kenneth L. Bauge: chairman, departments of business administration and economics, Augustana College, Sept. 1971.

Leland Burns: associate dean, School of Architecture and Urban Planning, University of California, Los Angeles.

Donald D. Christenson, University of Illinois: dean, College of Business Administration, Wichita State University.

Dan L. Costley, Michigan State University: dean, College of Business Administration, Wichita State University.

Vincent Davis, U.S. Naval War College: director of Patterson School of Diplomacy, University of Kentucky, August 1971.

William F. Ford, Transylvania University: director of research and forward planning, The American Bankers Association, Aug. 1, 1971.

Richard Hellman, U.S. Small Business Administration: professor of economics; director, Research Center in Business and Economics, University of Rhode Island.

Joseph J. Horton, Jr.: professor; chairman, department of economics, Slippery Rock State College, July 1971.

Robert S. Johnson, National Association of Mutual Savings Banks: dean, School of Business and Economics, California State College at Hayward, Sept. 1, 1971.

Charles H. C. Kao: chairman, department of economics, Wisconsin State University-River Falls, June 1971.

Irving Leveson: assistant administrator for Health Systems Planning, New York Health Services Administration.

John Lindauer: professor; chairman, department of economics, Claremont Men's College.

John W. McConnell: acting chairman, department of economics, Ithaca College, Sept. 1971.

James W. McKie, Vanderbilt University: dean, College of Social and Behavioral Sciences, University of Texas, Austin, Sept. 1, 1971.

Dimitri Mavros, Baruch College, City University of New York: executive vice president and director of research, Planning and Research Consultants, Ltd, Athens, Greece.

Herbert E. Newman: director of graduate studies, Hood College, Sept. 1971.

Martin M. Perline, Ohio State University: dean, College of Business Administration, Wichita State University.

Orme W. Phelps: dean of faculty, Claremont Men's College.

James E. Price: chairman, economics department, Syracuse University.

John Rafferty: acting chief, Economic Analysis Branch, National Center for Health Services Research, Maryland.

James W. Robinson, Virginia Polytechnic Institute and State University: head; professor; department of economics, University of Akron.

W. James Truitt: professor; chairman, department of economics and finance, Baylor University, Sept. 1971.

Leonard A. White: head, department of economics, College of Business Administration, University of Arkansas, July 1, 1971.

New Appointments

George T. Abed, University of California, Berkeley: assistant professor of economics, California State College at Hayward, fall 1971.

Keith A. L. Acheson: associate professor of economics, Carleton University, July 1, 1971.

Sidney Afriat, University of North Carolina: visiting professor, economics department, University of Waterloo, 1971–72.

Terry D. Aranoff: lecturer, department of economics, Rutgers-The State University, July 1, 1971.

Rachel Balbach: visiting assistant professor, department of economics, Washington University, 1971–72.

Wolde Bekele: economist foreign research division, Federal Reserve Bank of New York.

Richard A. Brecher, Harvard University: assistant professor of economics, Yale University.

Alan A. Brown, University of Southern California, Los Angeles: professor of economics, University of Windsor.

Robert E. Bruneau: instructor, department of mineral economics, Pennsylvania State University, Sept. 1971.

Merrill K. Bunker: department of economics, Brigham Young University.

David Butterfield, University of California: assistant professor of economics, McMaster University.

Alexander E. Cassuto, University of California, Los Angeles: assistant professor of economics, California State College at Hayward, fall 1971.

Samuel B. Chase, Jr., University of Montana: associate director, division of research and statistics, Board of Governors of the Federal Reserve System.

Chu-yuan Cheng: associate professor of economics, Ball State University, Sept. 1971.

Terrence M. Clauretie, Washington State University: assistant professor of economics, Marshall University.

Thomas Cody: economist, domestic research division, Federal Reserve Bank of New York.

Paul H. Cootner, Massachusetts Institute of Technology: professor of finance, Stanford Graduate School of Business.

Larry L. Cummings, University of Wisconsin: visiting professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

George G. Daly: associate professor, department of economics, University of Houston.

Joseph S. DeSalvo, The Rand Corporation: associate professor of economics, University of Wisconsin, Milwaukee.

John R. Dominguez: assistant professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

Arthur W. Donner: assistant professor of economics, McMaster University, 1971–72.

Arthur C. Eckerman: professor of industrial relations, department of business and economics, West Chester State College, Sept. 1971.

Bruce C. Fauman, Stanford University: visiting assistant professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

Gerald A. Feltham, Stanford University: associate professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

Kenneth C. Fraundorf: assistant professor of economics, University of Akron.

Melvin C. Fredlund, Wayne State University: assistant professor of economics, California State College at Hayward, fall 1971.

Herb Genfan: assistant professor of business administration, division of business administration, Ithaca College, Sept. 1971.

Lowell M. Glenn, National Manpower Policy Task Force: department of economics, Brigham Young University.

Steven D. Gold: acting instructor of economics, Drake University, Sept. 1971.

Eileen Gramm, University of Wisconsin: assistant professor, department of economics, University of Waterloo, 1971–72.

Harvey Gramm, University of Wisconsin: assistant professor, economics department, University of Waterloo, 1971–72.

Richard W. Grayston, University of Chicago: visiting assistant professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

Arthur L. Grey, University of Washington: visiting professor of economics, Claremont Men's College.

Arthur Guthrie: assistant professor of accounting, department of economics, Carleton University, July 1, 1971.

Donald L. Hooks, Texas A&M University: assistant professor, department of economics, University of Alabama.

George Horwich, Purdue University: visiting professor of economics, University of California, San Diego, 1971–72.

Lewis Kochin: economist, research and statistics function, Federal Reserve Bank of New York.

George A. Kraft: instructor of economics, Drake University, Sept. 1971.

Peter A. Lawrence: visiting assistant professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

Jae W. Lee: assistant professor of economics, Baruch College, City University of New York, 1971–72.

Maurice D. Levi, University of Chicago: postdoctoral fellow, faculty of commerce and business administration, University of British Columbia, Jan. 1972.

James T. Little: assistant professor of economics, Washington University, fall 1971.

Wesley H. Long, Pennsylvania State University: economist, U.S. Department of Commerce, Economic Development Administration.

James L. McCabe, Agency for International Development: assistant professor of economics, Yale University.

Thomas R. McKinnon: assistant professor of economics, College of Business Administration, University of Arkansas, Aug. 1971.

John C. McManus: assistant professor of economics, Carleton University, July 1, 1971.

Robert S. Main, University of California, Los Angeles: assistant professor of economics, California State College at Hayward, fall 1971.

Manoranjan Majumdar: visiting assistant professor, department of economics, University of New Mexico.

John D. Milliken, University of Kentucky: department of economics, College of Wooster.

William G. Moss, University of California, Berkeley: acting assistant professor, University of California, Davis, 1971–72.

Robert Mundell, University of Chicago: visiting professor, economics department, University of Waterloo, fall 1971.

James A. Ohlson, University of California, Berkeley: postdoctoral fellow, faculty of commerce and business administration, University of British Columbia, fall 1971.

Bruce M. Owen: senior economist, office of Telecommunications Policy, Executive Office of the President.

Robert P. Parks: assistant professor of economics, and economic research, Social Science Institute, Washington University, fall 1971.

Anthony Raia: associate professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

James P. Rakowski, Columbia University: postdoctoral fellow, faculty of commerce and business administration, University of British Columbia, fall 1971.

Andrew L. Robb, University of Essex: assistant professor of economics, McMaster University.

Refugio I. Rochin, Ford Foundation Pakistan program: acting assistant professor, University of California, 1971–72.

Walter Rountree, Virginia State College: assistant professor of economics, Wittenberg University, fall 1971.

Karl M. Ruppenthal, Stanford University: faculty of commerce and business administration, University of British Columbia, Jan. 1971.

George Rybakoff: associate professor of business administration, division of business administration, Ithaca College, Sept. 1971.

Jerry St. Dennis, University of California, Los Angeles: instructor of economics, Claremont Men's College.

William M. Scarth, University of Toronto: assistant professor of economics, McMaster University.

Jeffrey B. Sidney: assistant professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

Carl M. Smith: assistant professor of accounting, department of business and economics, West Chester State College, Sept. 1971.

Wallace F. Smith, University of California, Berkeley: professor, faculty of commerce and business administration, University of British Columbia, Jan. 1972.

Loren Solnick, Cornell University: assistant professor, department of economics, Claremont Men's College.

John C. Soper, Central Michigan University: visiting assistant professor, department of economics, University of Missouri-Columbia.

Thaddeus Spratlen: associate professor, Graduate School of Management, University of California, Los Angeles, July 1, 1971.

Mark Thompson, International Labour Office, Geneva: assistant professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

James W. C. Tomlinson: visiting associate professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

James T. Turner: instructor in economics, University of Akron.

Carl Van Duyne: instructor, economics department, U.S. Naval Academy, summer 1970.

Walter J. Wadycki, Miami University (Ohio): visiting lecturer in quantitative methods, University of Illinois at Chicago Circle, 1971–72.

James E. Walter, University of Pennsylvania: visiting professor in finance, faculty of commerce and business administration, University of British Columbia, fall 1971.

Robert Warne, University of Western Ontario: assistant professor of economics, McMaster University.

Peter D. Watts: assistant professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

Frederick S. Weaver, University of California, Santa Cruz: associate professor of economics, School of Social Science, Hampshire College, fall 1971.

Alma B. Weber: instructor of economics, University of Alabama.

Sidney Weintraub, University of Pennsylvania: consultant to economics department, University of Waterloo, 1971–72.

Doyle L. Weiss, Purdue University: visiting professor, faculty of commerce and business administration, University of British Columbia, fall 1971.

Robert W. White, Massachusetts Institute of Technology: assistant professor, faculty of commerce and business administration, University of British Columbia, Jan. 1972.

Leaves for Special Appointments

Irving Brecher, McGill University: senior fellow, East-West Center, University of Hawaii, Sept. 1971–

Mar. 1972; visiting fellow, Research School of Pacific Studies, Australian National University, Apr.–Sept. 1972.

Andrew C. Gross, Cleveland State University: visiting international research and exchange board scholar, Hungarian Institute for Market Research, Budapest, 1971–72.

Marshall M. Hall, Washington University: Makerere University, Kampala, Uganda, 1971–72.

John W. Hannaford, Ball State University: financial economist, U.S. Treasury Department.

Jerry R. Ladman, Arizona State University: program assistant in program management, Ford Foundation, Mexico City, 1971–72.

Robert Lindsay, Graduate School of Business Administration, New York University: executive director, State of New York Council of Economic Advisers.

David E. Lindsey, Ohio State University: visiting assistant professor, Macalester College.

David R. Miller, Syracuse University: economic policy fellow, Brookings Institution, 1971–72.

Jerry Miner, Syracuse University: research fellow, Higher Education Research Unit, London School of Economics and Political Science, 1971–72.

Raul Moncarz, Louisiana State University: visiting professor, Universidad CentroAmericana, El Salvador.

Richard D. Porter, Ohio State University: Board of Governors of Federal Reserve System, Washington.

Vinod Prakash, University of Rhode Island: economist, economics department, International Bank for Reconstruction and Development.

Tsung-Yuen Shen, University of California, Davis: University Development Program, Makerere University, Kampala, Uganda, 1971–72.

Allen Sinai, University of Illinois at Chicago Circle: senior economist, Data Resources, Inc., Lexington, Mass. 1971–72.

Robert E. Smith, University of Oregon: chief, Restrictive Business Practices Section, Manufacturers Division, UNCTAD, United Nations, Geneva.

Nat Weinberg, United Auto Workers: regent's professor, department of economics, University of California, Davis, Jan.–Apr. 1972.

Resignations

Gary Erickson, University of Akron.

Ulrich F. W. Ernst, Ohio State University: Abt Association, Inc., Cambridge, Mass., Sept. 1971.

Jo F. Freeman, Texas A & I University at Laredo, June 1971.

Hamid Habibagahi, University of Waterloo.

George Keyt, Federal Reserve Bank of New York: General Motors Corporation.

Richard L. Kozelka, interim president of United Theological Seminary of the Twin Cities, Sept. 1971.

Shumpei Kumon, Carleton University.

Anton S. Nissen, Federal Reserve Bank of New York: Board of Governors of the Federal Reserve System.

Kyprianos Prodromidis, Carleton University.

P. Foster Roden, Baylor University: Louisiana State University in New Orleans, Sept. 1971.

Sherman Shapiro, University of Illinois in Chicago: Federal Home Loan Bank, Chicago.

M. Richard Sussman, University of Georgia, Aug. 1971.

Burke O. Trafton, Pennsylvania State University, Sept. 20, 1971.

G. Richard Wynn, College of Wooster.

SIXTY-EIGHTH LIST OF DOCTORAL DISSERTATIONS IN POLITICAL ECONOMY IN AMERICAN UNIVERSITIES AND COLLEGES

The present list specifies doctoral degrees conferred during the academic year terminating June 1971. Abstracts of many of the dissertations are supplied.

General Economics; including Economic Theory, History of Thought, Methodology, Economic History, and Economic Systems

MOHAMMAD A. AKHTAR, Ph.D. South Carolina 1971. English classical macro-economic models.

The dissertation contains extensive discussions of some of the recent mathematizations of Ricardian economics and a new long-run model of Ricardian-classical macroeconomics. The latter is different from other Ricardian models in two respects. First, the new model substantiates the nonpessimistic interpretation of classical economics by incorporating fixed capital and discrete time to analyze the effects of technical progress. Second, the model includes a comprehensive treatment of classical monetary and aggregate demand theories.

AUSAF ALI, Ph.D. Southern California 1970. The political economy of the Islamic State: A comparative study.

The Islamic economic theory has certain fundamental postulates which it cannot violate. These postulates are the Quranic injunctions with regard to social justice. The most significant of these are (1) the categorical abolition of interest, (2) priority of community interest over the individual's, and (3) responsibility of the state to provide social and economic security to all members of the society. Islam envisions a welfare state which would be somewhere in the middle between capitalism and socialism.

ANASTASIOS ANASTASOPOULOS, Ph.D. Rochester 1971. A study on the theory of uncertainty.

Part I: The optimum mix of financial assets. Topics: Portfolio choice, prices of assets in equilibrium, the price of risk. Part II: The loan market with risk of default. Topics: The demand and supply of loans under conditions of risk of default, existence and stability of equilibrium. Part III: Consumer behavior over time under price uncertainity. Topics: The utility of income streams, optimum consumption policy, the effect of a change in price expectations on optimizing decisions.

HUGH T. ARTHUR, Ph.D. South Carolina 1971. A model of capital employment.

This dissertation is a static, aggregate, abstract model of the markets for the employment of the factors of production. The two markets are those for the given stocks of human capital and physical capital. These two classes of capital goods are treated as heterogeneous variables in the aggregate production function. Thus, employment is treated completely within the framework of capital theory. The uniqueness of the model stems from the hetergeneous human capital variable.

TRUMAN BEWLEY, Ph.D. California (Berkeley) 1970. Equilibrium theory with an infinite dimensional commodity space.

MILTON N. BUECHNER, Ph.D. Virginia 1971. A critical exposition of Frank H. Knight's conception of capital and the production process.

Knight's unusual view of the production process depends completely on his conception of capital as a supernatural value substance. Given this conception of capital, one can then proceed to make intelligible his propositions that "production and consumption are simultaneous" and "capital is permanent." Since this concept is a pure mystical invention for which there is not a single shred of evidence, it is entirely invalid, and Knight's whole structure of analysis collapses.

HENRY S. BURNESS, Ph.D. Kansas 1970. Impatience: programs with finite and infinite horizons.

The thesis is an attempt to delineate the relationship between separability and existence of functions defined over infinite dimensional horizons and the presence of impatience in the sense of Koopmans (*Econometrica*, 1960). Also some theorems on finite dimensional spaces are presented, and the effect of goods that adversely affect the lifespan of individuals is investigated, much after the method of Yaari (*Review of Economic Studies*, 1965).

MICHAEL J. CHOSSUDOVSKY, Ph.D. North Carolina 1971. Social welfare function to the planning of economic and social policy.

JOSEPH R. COWART, Ph.D. Duke 1971. The economic doctrines of Amasa Walker.

GARRY M. CRANE, Ph.D. George Washington 1971. Economic effects of an improved measurement: a micro-economic approach to a technological change.

Three issues related to improved measurement are studied. First, Lancaster's recent work is developed to show a theoretical method of analyzing the economic effects of an improved measurement. Second, a case study of an improved measurement used in manufacturing stainless steel is shown to have provided annual savings of several millions of dollars. Finally, it is shown that the private market is, generally, the preferred supplier of measurements notwithstanding their many public-goods characteristics.

THIERRY DE MONTBRIAL, Ph.D. California (Berkeley) 1971. Intertemporal general equilibrium and interest rate theory.

ALLAN C. DE SERPA, Ph.D. California (Santa Barbara) 1970. A general theory of the economics of time.

This study addresses the consumer's problem of concurrently allocating money and time on common competing uses. Utility is assumed a function of commodities and amounts of time, and time and money resource constraints are distinct. Most importantly, the physical relationship between the quantity demanded of commodities and the amount of time allocated to them is made explicit as an additional set of constraints. This generates equilibrium values of saving time from each consumption activity.

GERALD J. ESKIN, Ph.D. Minnesota 1971. Prediction of demand under a change in consumption technology.

A utility maximization model of the Lancaster latent structure type is specified in order to investigate problems related to prediction under a form of structural change. The demand relations for a quadrative representation of utility are derived. Integrability conditions on the demand relation for the existence and uniqueness of a utility indicator are established. Conditions for the existence and uniqueness of the latent structure are also determined.

MARY L. EYSENBACH, Ph.D. Stanford 1970. American manufactured exports, 1879–1914: a study of growth and comparative advantage.

ROBERT T. FALCONER, Ph.D. Texas A&M 1971. Capital intensity and the real wage: a critical evaluation of Hayek's Ricardo Effect.

The scope of this dissertation is an exposition and criticism of the theoretical aspects of the Ricardo Effect as developed by Frederich von Hayek. First, the general background necessary for understanding the development of the Ricardo Effect is presented. The second part examines Hayek's contention that Ricardo developed the Ricardo Effect. The third part presents Hayek's various interpretations of the Ricardo Effect. The final section reviews the theoretical and empirical criticisms of the Ricardo Effect.

WILLIAM B. GINSBERG, Ph.D. Harvard 1971. Nonconvexities in production.

Traditional neoclassical assumptions of convexity are weakened in an attempt to extend the partial equilibrium analysis of the profit-maximizing competitive firm to special cases of indivisibilities and increasing returns to scale. A special class of "convex-concave" functions is defined and their properties examined. These are used in a study of activity analysis. The optimal allocation of activities is a theoretically difficult problem and does not lend itself to a general solution.

JEAN-MICHEL GRANDMONT, Ph.D. California (Berkeley) 1971. On the temporary competitive equilibrium.

JAMES M. GRIFFIN, Ph.D. Pennsylvania 1971. Economic capacity in the joint product case: the *U.S.* petroleum refining industry.

Capacity is defined as the minimum of the industry's short-run average cost function. A process analysis approach is used to derive capacity, which is measured along a fixed product mix. The variable product mix case is also examined along with the effects of relative product price changes on capacity. Finally, the economic capacity measure is compared to the Wharton capacity measure for petroleum refining and to the industry's standard engineering measure.

BOWMAN N. HALL II, Ph.D. Duke 1971. A history and critique of American individualist anarchists' economic theories.

JOHN M. HARRINGTON, JR., Ph.D. Georgetown 1971. General equlibrium and regional allocation of money resources: a methodology for implementing general equilibrium analysis.

Scarf's algorithm for the computation of equilibrium prices is extended to admit primitive sets from any appropriately bounded subset of any hyperplane parallel to the unit simplex. This algorithm is then used to determine the optimum allocation of money resources to a region so that the region will be in equilibrium with the economy in which it is embedded. Public goods are introduced into the general equilibrium framework and some hypothetical numerical examples are presented.

WALTER P. HELLER, Ph.D. Stanford 1970. Disequilibrium dynamics of competitive growth paths.

Disequilibrium behavior in an infinite-horizon Ramsey model is examined. Firms profit-maximize and households maximize utility. Analogous to the standard tatonnement analysis, consumption goods prices rise if there is excess demand, or interest rates rise if there is excess demand for capital. Under some boundedness assumptions, it is shown that the long-run equilibrium path is stable. Thus, there is a disequilibrium process which beats the saddle-point problem.

DONALD L. HOOKS, Ph.D. Texas A&M 1971. The Wicksell Effect and its implication for modern capital theory.

The current controversy regarding the generality and even the viability of the neoclassical theory of distribution and capital often makes reference to "Wicksell Effects." One aspect of the current controversy, capital revaluation, may be explained in terms of price or financial Wicksell Effects. A second criticism of neoclassical theory involves the reswitching of techniques, and it may be explained in terms of real Wicksell Effects. The various effects merely explain the workings of a neoclassical model with no a priori reason for designating them as "perverse."

JOE R. HULETT, Ph.D. Iowa State 1970. A theory of jointness and infeasibility of exclusion.

This study investigates the theoretical relation between public goods and externalities. The allocative significance of jointness is examined for goods characterized by infeasibility of exclusion. Also analyzed are various suggested approaches to the optimal provision of goods characterized both by Marshallian joint supply and by jointness.

YUTAKA IMAI, Ph.D. Rice 1971. An integration of neoclassical theory of optimal capital accumulation and corporate financial theories.

In the thesis the neoclassical theory of optimal capital accumulation and the theories of corporate finance are integrated into a theory of a publicly owned firm operating under certainity. A formal correspondence between the Modigliani-Miller cost of capital theorems and Jorgenson's investment theory is established. Also, the relationship between the two longstanding controversies in corporate finance, i.e., whether leverage influences the cost of capital, and whether dividend policy affects a share price, are clarified.

BERNARD KEY, Ph.D. California (Berkeley) 1971. The role of foreign contributions in Japanese capital formation, 1868–1936, with special reference to the period 1904–14.

GEORGE LERMER, Ph.D. McGill 1971. Dynamic analysis of the Walrasian tatonnement.

This dissertation makes the following behavioral assumptions: firms face real adjustment costs that are a monotone increasing function of the speed at which they adjust; consumers attach a disutility to adjustment that is a monotone increasing function of the speed of adjustment of their consumption basket. These assumptions permit the problem of the speed of adjustment of both economic units and economic systems to an exogenous change to be investigated by means of an optimizing model.

LOUIS J. MACCINI, Ph.D. Northwestern 1970. Delivery lags and the theory of investment.

STUART MESTELMAN, Ph.D. Purdue 1970. An investigation of long-run static equilibrium properties in situations affording external diseconomies.

First, the effect of the assignment of liability on resource allocation is examined in a long-run partial equilibrium setting. Next, a general equilibrium situation is developed to ascertain whether or not subsidies and taxes could result in an economy's achieving a given point on its production possibilities frontier in the long-run. Finally, a general equilibrium welfare framework is introduced for the purpose of examining the use of any tax receipts that a governmental regulatory agency may accumulate if a tax scheme is chosen as a corrective device.

JOHN W. MITCHELL, Ph.D. Oregon 1970. New South Wales 1800–1900: a case study in the Staple theory.

LAURENCE S. MOSS, Ph.D. Columbia 1971. Mountifort Longfield's contribution to economic theory and its place in the development of British economic thought.

Longfield's theory of distribution may properly be termed "non-Ricardian" because it is based on an original theory of profit that explicitly denies that the application of capital to land plays any special role in regulating the return on invested capital throughout the economy. Of Ricardo's many critics, Longfield appears to have been the only one to have developed an alternative theoretical system which had some influence on English, Irish, and Australian economic thought.

DANIEL H. NEWLON, Ph.D. Virginia 1970. Income redistribution as investment and insurance.

HIROSHI OHTA, Ph.D. Texas A&M 1971. Aspects of spatial pricing.

This dissertation endeavors to investigate the impacts of economic space and spatial competition on the pricing practice and outputs of the firm. Our analysis yields conclusions significantly different from those of classical spaceless economics. In particular, this dissertation sets forth several propositions, namely, (1) spatial monopolists necessarily tend to practice discriminatory pricing, (2) spatial competitors will often discriminate in price over the space, and (3) spatial discrimination necessarily yields more outputs than nondiscriminatory pricing.

ALAN L. OLMSTEAD, Ph.D. Wisconsin 1970. New York City mutual savings banks in the antebellum years: 1819–61.

This dissertation examines whether the behavior of early New York mutual savings banks was consistent with the stated philanthropic goals of their trustees. Some of the topics studied are changes in administrative framework, the growth of competition among mutuals, the socioeconomic status of depositors, the impact of legal constraints on mutual behavior, the role of mutuals in financing antebellum development, and the determinants of mutual investment policy.

ERVOL O. OLSEN, JR., Ph.D. Texas A&M 1970. Utility maximization and the theory of the firm.

The behavior of a utility-maximizing individual who owns and manages his own firm is the center of analysis. Since his income is determined by the behavior of his firm, his labor-leisure choice determines the income he will have available for purchasing commodities, which, like leisure, yield positive utility. It is shown that in long-run competitive equilibrium the owner manager must maximize the profit of his firm in order to maximize his utility.

MICHAEL J. PANIK, Ph.D. Boston College 1970. Specification problems in the derivation of *CES* production functions.

ROBERT P. PARKS, Ph.D. Purdue 1971. Rationalization, extension, and social choice paths.

MICHAEL A. PERELMAN, Ph.D. California (Berkeley) 1971. Investment theory in light of expectations.

A firm's decision to invest is made by comparing the expected return from its existing capital with that of alternative pieces of capital, but the expected return from any potential unit of capital depends in part upon when it is replaced and that replacement in turn depends on future replacement decisions, etc. Thus a determination of the optimal decision can only be made in light of all future investment decisions. Mathematical models of this effect are made, and it is shown how the effect of the future can be collapsed into a measure of Keynesian animal spirits. The model is further modified to apply it to tractor investment.

GEOFFREY C. PHILPOTTS, Ph.D. Cornell 1971. Vote trading in a competitive market.

MICHAEL S. PROCTOR, Ph.D. Texas A&M 1970. Investment theory for the firm: deterministic and stochastic models.

WILLIAM F. RENTZ, Ph.D. Rochester 1971. Optimal consumption and portfolio policies.

This essay uses a discrete time-infinite horizon dynamic programming model of a representative individual who derives utility from consumption and a bequest. The individual's resources consist of an initial level of wealth and a salary income stream, and his investment opportunities consist of a risk-free asset and a finite number of risky assets including term insurance. Optimal consumption and portfolio policies are found for the Pratt-Arrow constant relative risk-averse functions.

JAMES D. RODGERS, Ph.D. Virginia 1970. Utility interdependence and income redistribution.

TIMOTHY P. ROTH, Ph.D. Texas A&M 1970. A synthesis and critique of the search theory literature with special reference to its application to the detection of alternative production techniques.

On the assumption that a product may be produced with the use of any one of a number of alternative production "techniques," a multi-equation production function is posited. Each subjunction or technique is associated with a unique form of the capital input. Because the number of techniques may be very large a search problem exists. The optimal distribution of search effort is derived both for the case of an industry entrant confronted with an array of contemporaneously available techniques, and the case in which techniques became available sequentially over time.

ABRAHAM J. SIMON, Ph.D. Pennsylvania 1971. The income function—the consumption function existence and interdependence at the micro level, investigations into specification difficulties in a market and non-market setting.

The current study is directed toward the position that associated with any consumption function, at the level of the individual household, is an actual or potential income function. The usual formulation of the income-consumption relation at the micro level for purposes of empirical research is that consumption is a function of income. We examine the case for the possible and significant existence of the converse relation that income is a function of consumption at the micro level.

JOHN C. SOPER, Ph.D. Massachusetts 1970. The long swing in historical perspective: an interpretive study.

MARTIN C. SPECHLER, Ph.D. Harvard 1971. The economics of product quality in Soviet industry.

Familiar irrationalities of planning and incentives under bureaucratic socialism do not account for demonstrable improvements of many Soviet goods. While reforming prices and contract law, Soviets have developed substitutes for market devices—standards, certification, buyer-seller ties, intra-plant organization. The pattern of quality success demonstrates the historical and institutional problems. Past priorities have weakened the capabilities of the Soviet enterprise to turn out an optimal quality. Inflation and distrust still hinder both market and substitute mechanisms.

DARIO TOSI, Ph.D. Warsaw 1971. External economics in production—an essay on foundations.

KAREN I. VAUGHN, Ph.D. Duke 1971. An examination of the economic theories of John Locke.

GOTTFRIED E. VOELKER, Ph.D. Nebraska 1970. The impact of railroads on the German economy in the late nineteenth century.

After a review of the early railroad history in Germany, the railroads' impact on the economy is developed in two steps: the portion that is quantitatively measurable and the portion that is qualitatively describable. The quantitative portion is measured in social saving, while the qualitative portion is described through the railroads' traffic-significance and formative power. In conclusion, it was found that the impact of the railroads on the development of the economy was highly important.

PETER VON ZUR MUEHLEN, Ph.D. Princeton 1970. Optimal price decisions over time.

Given the existence of restrictions on adjustment of price, output and inventories, several theoretical monopoly and oligopoly models of price determination are developed assuming maximization of a (net) revenue functional. The principal conclusions are (1) the timing of optimal price decisions is endogenously determined, (2) permanent and transitory phenomena are interdependent, and (3) the turnpike of the solution involves the forward histories of cost and demand determinants. For a quadratic revenue approximation this is empirically verified using manufacturing data.

RONALD W. WARD, Ph.D. Iowa State 1970. Some theoretical considerations for futures trading in commodities requiring transformation services: the case for live beef futures.

The thesis develops a theoretical model of optimal firm decisions in cash and futures markets that includes both primary product producers and marketing firms. The generalized model of production and marketing decisions under risk was applied to both short and long hedging and speculation. Hedging and speculation were given precise definitions. Live beef futures were used to show how transformation costs for nonstorable commodities should be treated in the same manner as storage costs for storable commodities.

Donald A. Wittman, Ph.D. California (Berkeley) 1970. Theories of optimal party decision making.

Joseph Shao-Kong Wu, Ph.D. Purdue 1971. Resource allocation in a nonconvex economy.

In this thesis we use a geometrical construction based on the conjugate function theory to illuminate the relationship between Pareto optima, equilibria and net social welfare maxima, in an economy where production set is not necessarily convex. We formulated the characterization of the necessary and sufficient conditions for a Pareto optimal allocation to be attainable as a net social welfare maximum decentralizable into a competitive equilibrium, in a nonconvex production economy.

Stephen H. Zeller, Ph.D. Boston College 1971. The urban firm's production-location decision: towards a theory of land and property taxation.

The profit-maximization model is based on a Cobb-Douglas produoctin function in land, labor, and capital. Prices are exogenous and the firm locates along a straight line. Solutions are obtained with computer-aided numerical analysis techniques. The long-run comparative-static behavior of the firm is developed and it is seen that the imposition of either a land or a property tax will cause the firm to reduce its equilibrium stock of capital.

Economic Growth and Development; including Economic Planning Theory and Policy, Economic Fluctuations and Forecasting

Faik A. Abdul-Rasool, Ph.D. Massachusetts 1970. The effect of the pattern use of oil revenues on the growth and prices of Iraq.

Curtis H. Adams, Ph.D. Iowa 1971. Investment decision making for development: the Singer proposal.

Mohiuddin Alamgir, Ph.D. Harvard 1971. A planning model for East Pakistan with special emphasis on manpower and education.

Magdaleno B. Albarracin, D.B.A. Harvard 1971. Plant capacity decision making in the Philippine cement industry.

This thesis examines the process by which Philippine cement firms made their capacity decisions and suggests concepts and techniques for improving such decisions. Thesis research involved field interviews with managers of cement firms, management consultants, officials of government financial institutions, key personnel of the Cement Association of the Philippines, and computer company executives. It includes a process note on cement manufacturing.

James E. Annable, Jr., Ph.D. Princeton 1971. The transfer of technology, industrialization and the urban employment problem in low-income countries

The study has three parts: 1) The internationa transfer of technology is used as the framework within which to analyze the components of the urban employment problem in *LDC*'s: low industrial labor-absorption capacity, high population growth, and high rates o rural out-migration. 2) I derive a three-sector model o labor-force distribution in the *LDC*'s, and deduce equilibrium rates of urban unemployment, optimal labor force distribution and the time-path of urban unemployment. 3) The model's predictions are tested.

Ghassan M. Arnaoot, Ph.D. Texas (Austin) 1970. The role of foreign trade in economic development Syria, a case study.

This dissertation examines the role of foreign trade in the economic development of Syria. Several models were utilized to explain the foreign-exchange gap problem during 1957–68. The study demonstrates that the government development and investment policies were the main causes of the foreign exchange gap.

Harry W. Ayer, Ph.D. Purdue 1970. The costs, returns and effects of agricultural research in a developing country: the case of cotton seed research in São Paulo, Brazil.

The aim of this study was to add empirical knowledge concerning agricultural research as a source of economic growth in developing nations. It focuses on the cotton seed research and development program of São Paulo Brazil. The costs and returns to society from investments in cotton seed research were estimated. The internal rate of return was at least 95 percent. Program effects on cotton fiber exports and on Brazilian cotton cloth consumers were evaluated.

Donald E. Baer, Ph.D. Illinois (Urbana) 1971 Taxation and the economic development of Costa Rica and Honduras with emphasis upon consumption and agricultural based taxes.

The dissertation initially analyzes the implementation of retail sales taxation and consumption excise-type taxation in two developing countries: Costa Rica and Honduras. Conclusions are based upon original aggregations of individual tax declarations of the two countries The second emphasis of the study is an analysis and evaluation of the significance of the taxation of the agricultural sector in the two countries.

Theophanis Benos, Ph.D. Rochester 1971. An empirical two-sector model for the *U.S.* economy and the effects of alternative taxation policies.

The thesis examines two problems. The first deals with the derivation of the actual growth path followed by the *U.S.* economy during the period 1947–67, on the basis of observed savings decisions by *U.S.* consumers. In this respect a two-sector two-commodity general equilibrium model is used with Cobb-Douglas production functions incorporated in both sectors. The second problem deals with the differential impact of alternative government taxing policies on the rate of capital formation, as well as on the other endogenous variables of the system, under the assumption of the government's balanced budget.

ELIZABETH C. BOGAN, Ph.D. Columbia 1971. Investment plans and their realization.

BASIL A. BROWNE, Ph.D. Washington State 1971. An evaluation of the economic development ideas in the reports of the World Bank's general survey missions.

This dissertation examines the published reports of the first twenty-five General Survey Missions of the World Bank in order to discover their underlying economic rationale and to identify the ideas in development theory to which this rationale corresponds, focusing on agriculture, manufacturing and trade, and transportation. It is found that the Missions usually placed principal emphasis on improving agriculture.

RICHARD A. CALKINS, Ph.D. Duke 1970. Technology, innovation, and economic development: The Indian industry experience.

SE-MOON CHANG, Ph.D. Florida State 1971. Transformation of dual economies with special reference to internal migration and marginal product of farm labor.

The estimates of our model of internal migration which is based on the supply and demand relation indicate that education and the demand side of migration are more important than wage differential in explaining the behavior of internal migration of Korea, Japan, and Taiwan. Our estimation also indicates that the average ratios of the *MRP* of farm labor to those of nonfarm labor during 1952 to 1967 are .4151, .2618, and .2573 for Korea, Japan, and Taiwan. With the exception of Taiwan, the wage rates are larger than the *MRP* of labor, but smaller than the average products in both farm and nonfarm sectors.

GHULUM R. CHAUDHRY, Ph.D. Kansas State 1970. Marketing institutions and economic development.

The study deals with some of the newly emerging marketing institutions of *LDC*'s especially introduced and developed after World War II. The institutions included in this study are Marketing Boards, Government Regulated Markets, Cooperatives, International Agreements. An attempt has been made to describe the working of these institutions and to analyze and appraise their efficiency. Effort has also been made to trace out the objectives for which these new institutions were needed.

PHILLIP E. CHURCH, Ph.D. Oregon 1970. Traditional agricultural markets in Guatemala: their structure and performance in a developing economy.

THOMAS W. CLEAVER, Ph.D. Harvard 1971. Regional income differentials in Japanese economic growth.

This study compares original per capita income estimates for forty-six Japanese prefectures in 1890, 1905, and 1920 to modified existing estimates for 1930, 1935, 1940, and 1958-65. The pattern of inequality, an "inverted U," resembles experience found elsewhere. The early rise (1890-1905) is attributed to the breakdown of regional self-sufficiency as secondary industry, especially food processing, became more concentrated. The long-term fall since 1905 has been due mainly to the equilibrating effects of internal migration.

JOHN A. DADSON, Ph.D. Harvard 1971. Collective and state farms in Ghana's agricultural development.

Ghana's attempt to use socialized agriculture as a modernizing device is examined in terms of the initial preparations preceding its establishment: the acquisition and utilization of land; the production and distribution of commodities; the transfer of technology; the working arrangements governing farm resource use, in particular management and labor; and the organization administration and control of the sector. It is suggested that the program of technical modernization and socialist transformation was frustrated by lack of planning, poor implementation, and dysfunctional politics.

CARLETON G. DAVIS, Ph.D. Michigan State 1970. The organization and economics of agricultural research in the West Indies: strategy implications for the 1970's.

This thesis describes the present agricultural research system, analyzes the performance of the present research system and makes recommendations to improve the administration of research and determination of research priorities in the 1970's. The two major strategies recommended are: 1) That high priorities be given to livestock, cereals, fruits, vegetables and root crops; and that 2) research efforts be concentrated on generated bio-mechanical innovations of the 1970's with emphasis on sugar cane and bananas.

ROBERT M. DEAVER, Ph.D. Oklahoma 1971. An empirical examination of W. W. Rostow's historical stages of economic development.

Following a compilation of the literature surrounding Rostow's historical stages theory, United Nations time-series data were analyzed by regression analysis. The evidence indicated investment and government sectors generally occupying larger percentages of nations' economies as they became more developed. Some support was found for Rostow's hypothesis of a doubling of investment proportions during the take-off stage.

THEODORE E. DOTY, Ph.D. Washington State 1971. Washington State University in West Pakistan, 1954–69: an evaluation of technical assistance to higher education for agricultural and economic development.

The purpose of the study is to provide a conceptual and empirical base upon which to improve the effectiveness of technical assistance project implementation. Through an assessment of fifteen years of technical assistance in West Pakistan, provided by the Agency for International Development through a contract with Washington State University, the study confronts the need for more effective means of evaluating technical assistance from the *ex post* perspective.

PETER ECKSTEIN, Ph.D. Harvard 1971. Latin American development performance: a two-gap analysis.

SUNDAY M. ESSANG, Ph.D. Michigan State 1970. Determinants of income distribution among cocoa farmers and buyers in Western Nigeria: implications for food production, consumption, and rural development strategy.

This dissertation investigates empirically the influence of marketing board operations on the distribution of incomes among cocoa producers in licensed buying agencies in Western Nigeria. It also explored the possible implications of reducing marketing board "surpluses" for the distribution of incomes among the same group of people.

MANSOUR FARSAD, Ph.D. New York 1971. The role of capital formation in economic development with special reference to Iran.

This paper deals with the theoretical views and Iran's experience concerning the role of human and physical capital formation in development. First, based on an index of educational development, the rate of human capital formation is measured. Then, the statistical relationships between this index and other indicators of economic development is determined. Second, based on the assessment of methods of integrating educational and general development planning, an appropriate technique for integration of educational planning and development planning in Iran is proposed.

KWAME D. FORDWOR, Ph.D. Pennsylvania 1971. An evaluation of Ghanaian state-owned industrial projects: a study in cost-benefit analysis.

ALFRED M. FUERST, Ph.D. New York 1970. Economic development with special attention to the role played by human capital: the cases of Israel and Puerto Rico.

The basic questions analyzed are what makes a country develop and how does development manifest itself. Major emphasis is placed upon "diffusion" of development. Human capital emerges as the prime mover of economic development under different circumstances obtaining in Israel and Puerto Rico. It is also argued that "development" is measured by a combination of indicators representing economic growth and "income" to human capital as indicated by an index for "average nonpecuniary income."

RICHARD L. GADY, Ph.D. Purdue 1971. An analysis of the effect of inflation, government programs and technology on farm firm growth and adjustment.

The objective is an analytical framework incorporating exogenous environmental and firm decision variables to determine optimal firm growth patterns. Three situations are synthesized. A sequential multiperiod recursive programming model analyzed the effects of government programs, technological innovation, and inflation. Higher rate of inflation retarded growth in a physical sense, discouraged land ownership, and increased use of labor intensive facilities. Faster techn
logical innovation increased growth and use of capit
intensive facilities. Farms complying with governme
programs were able to grow at faster rates.

F. DERMOT GATELY, Ph.D. Princeton 1971. Investme
planning for the electric power industry: a mixe
integer programming approach, with application
southern India.

Questions of plant type, size, location, and tim
phasing are examined in an extension of the Électrici
de France linear programming framework. Zero-or
variables represent construction decisions for vario
proposed hydro sites and also scale economies in co
structing thermal and nuclear plants. Continuous var
ables represent thermal and nuclear plant size and al
plant output levels and interstate transmission.
branch-and-bound algorithm (Davis-Kendrick-Weit
man) is used. Questions of achieving regional cooper
tion are also examined in a game theoretic framework.

GARY G. GILBERT, Ph.D. Iowa 1971. Multicountr
planning for economic development: Latin Americ

NANCY M. GORDON, Ph.D. Stanford 1970. *Ex ante* an
ex post substitutability in economic growth.

To date, vintage growth models assume the existen
of labor-capital substitutability before specific machi
types are built and either the same degree or none aft
capital is in place. This paper allows some *ex post* su
stitution possibilities, although less than *ex ante*. A s
of necessary conditions for a steady state to exist a
derived and extensive computations are undertaken
demonstrate that the economy converges asympto
ically geometrically for wide ranges of paramet
values.

SARA L. GORDON, Ph.D. Stanford 1971. Aspects
economic development in Ghana.

This study investigated the impact upon the Ghan
ian economy of large expansions in the value of cocoa
determine to what extent growth in a primary expo
commodity can be transmitted to other sectors of t
economy. By examining the nature of linkages produce
and the consequent net gains to the economy, it w
shown that the noncomplex nature of cocoa's productio
function combined with institutional factors prevente
cocoa from effectively transmitting economic growth.

DAVID A. G. GREEN, Ph.D. Michigan State 197
Agricultural mechanization in Ethiopia: an econom
analysis of four case studies.

Four case studies of agricultural mechanization
Ethiopia are presented. The objectives are to analy
production and financial data with benefit-cost tec
niques over a ten-year period; to relate nonfinanci
aspects of these cases to Ethiopian development obje
tives; and to demonstrate the types of data required f
economic analysis of agricultural mechanization.

WALTER W. HAESSEL, Ph.D. Iowa State 1970.
theoretical analysis of intersectoral relationships in
five-sector optimizing model of a dual economy.

A five-sector model is used to investigate the relationships between the agricultural and nonagricultural sectors in the process of development. The five sectors include traditional agriculture, commercial agriculture, manufactured goods, capital goods, and government. It is concluded that in the earlier stages of the development of a "labor-surplus" economy, the expansion of agriculture would be desirable, but not to the exclusion of industrial development. As development proceeds, the emphasis should shift to industry.

JOHN R. HANSEN, Ph.D. Colorado 1970. The acquisition of technology for development.

VIRDEN L. HARRISON, Ph.D. Purdue 1970. Management strategies and decision processes for the growth of farm firms.

Growth in terms of income-producing potential was examined for a hypothetical cornbelt farmer whose complete operation was simulated over a fifteen-year period. The strategies tested involved purchase and/or rental of land, size of land parcel acquired, the down payment level and the length of repayment term, policy toward remortgaging land, and the use of livestock. Variations were considered in price and yield predicting ability and the starting asset position.

BENJAMIN G. HARTZOG, JR., Ph.D. Alabama 1970. An optimal economic stabilization mix: a synergistic approach.

KICHIRO HAYASHI, Ph.D. Indiana 1970. Technical change in Japan: its measurement, 1910–1967.

I attempt to clarify a conceptual framework to measure technical change aggregatively and to apply it to Japanese data. The results obtained for Japan are compared with those of Canada and the United States, with an emphasis on explaining any major differences between the countries. An attempt is also made to explain the Japanese residual.

TERRY HOGAN, Ph.D. California (Berkeley) 1970. A comparison of information structures and convergence properties of several multisector economic planning procedures.

SØREN J. HOLM, Ph.D. California (Berkeley) 1970. Informationally sufficient adjustment processes for an economy with externalities.

ROBERT C-M HSU, Ph.D. California (Berkeley) 1970. A study of the price of fertilizer in Taiwan, 1950–1966, first-best and second-best policies of taxation on traded input.

YUKON HUANG, Ph.D. Princeton 1971. The economics of paddy production in Malaya: an economy in transition.

This thesis studies how the introduction of the new technology associated with the introduction of double-cropping changes the behavior of subsistence farmers. A model of the subsistence paddy economy is formulated and tested using data from three paddy regions, each differing in the stage of development. Also considered are problems of land tenancy and its effect on productivity and socioeconomic problems associated with the introduction of double cropping.

KHALID IKRAM, Ph.D. Harvard 1971. Export policy in a developing country: a case study of Pakistan.

PETER ISARD, Ph.D. Stanford 1971. The development assistance content and domestic employment implications of *U.S.* textile imports: towards a plan for gradual contraction of the *U.S.* textile sector.

The untied aid grant which offers *LDC*'s the same development assistance as a unit of *U.S.* textile imports is formulated and estimated under different *LDC* planning objectives and bottlenecks to growth. Estimates of embodied and disembodied technical progress and projections of textile demand then lead to estimates of jointly feasible future time paths of *U.S.* textile employment, imports, and investment. The import streams are translated into equibeneficial aid streams.

DILMUS D. JAMES, Ph.D. Michigan State 1970. The economic feasibility of employing used machinery in less developed countries.

The advantages and disadvantages of employing used machinery in less developed countries are examined. Technical aspects of maintenance, learning and flexibility are considered as well as the institutions influencing the transfer of used machinery to developing countries. The study finds that the usefulness of second-hand equipment to developing countries is due to a variety of special situations rather than any predominant factor. The encouragment of a judicious expansion of the employment of used machinery appears to be a worthwhile policy objective for less developed countries.

KENNETH P. JAMESON, Ph.D. Wisconsin 1970. Adjustment costs and investment in the less developed country.

The thesis investigates the determinants of investment demand in India. The cost of adjustment model is used, adapted to account for Indian taxation and institutions. The model is fitted to data for twenty-five manufacturing industries. The process yields information both on the speed of adjustment in Indian industry and on the effects of taxation on investment.

JOHN F. JONES, JR., Ph.D. Florida 1970. Tourism as a tool for economic development with specific reference to the countries of Jamaica, Trinidad, and Guyana.

ANWAR N. KACYRA, Ph.D. Catholic 1970. The relation between size and efficiency in the developing enterprise economy of Iraq.

Two tests of efficiency were made, using industrial profitability and unit costs. On profitability, I investigated the pre-tax rate of return on capital for 1964–66 for 1,095 Iraqi enterprises. Profit rates generally increased with asset size up to I.D. 200,000. Beyond that point the relationship is uncertain. In the second test I reviewed unit production costs for the year 1966 reported by 194 Iraqi enterprises. In most cases, small scale enterprises show average unit costs higher than those of medium or large scale enterprises.

C. L. KENDALL, D.B.A. Harvard 1971. Consumer aspirations and change in Peruvian credit union members.

YAKUP KEPENEK, Ph.D. New York 1970. Investment-priority and economic development with specific reference to Turkey.

The study examines ways of changing the socio-economic structure in underdeveloped countries, with particular reference to investment in Turkey. A five-sector model (two capital goods and three consumption goods) is constructed with the specification of constraints on the growth of per capita consumption. Alternative investment allocation programs are examined from the viewpoint of time required to eliminate unemployment. The thesis demonstrates that investment should be allocated to produce capital goods used to make more capital goods, and the existing 20 percent unemployment can be eliminated with fifteen production periods.

PETER T. KNIGHT, Ph.D. Stanford 1970. Export expansion, import substitution and technological change in Brazilian agriculture: the case of Rio Grande do Sul.

This analysis of Brazilian agricultural development in the postwar period and its impact on that country's balance of payments focuses on the economics and related technologies of wheat, soybeans, rice, corn, and beef in the state of Rio Grande do Sul. Stagnation of land and herd productivity are documented. Attention is given to import substitution in wheat, export performance for the other four commodities, and the economics of fertilizer use.

TATSUO KOBAYASHI, Ph.D. California (Berkeley) 1970. Land reform in Japan.

The thesis deals with the land reform introduced in postwar Japan, with main emphasis on its effects on agricultural productivity. The basic proposition supporting the productivity objective of the land reform was that the reform contributed to increased productivity and output. The thesis controverts this proposition. Finally, the thesis points out that the postwar reform has obstructed general economic growth. An analysis indicates that the land reform has impeded an economic use of farm machinery, prerequisite to further economic development.

MANTHRI S. KRISHNASWAMY, Ph.D. Kansas State 1970. An analysis of the five-year plans in the agricultural development of Andhra Pradesh—India.

CARLOS G. LANGONI, Ph.D. Chicago 1970. A study in economic growth: the Brazilian case.

Estimates are made of the contribution of constant quality physical capital, raw labor, education, the age-sex component, and labor reallocation from low to high productivity sectors, to the Brazilian rate of growth during the 1950–69 period. Estimates are also made of the social rentability of investments in physical capital which are compared with that of human capital. The main conclusion is that there are important gains to be made without any change in the overall saving ratio, by shifting the emphasis from physical toward human capital.

TAKYUN J. LEE, Ph.D. State University of New York (Binghamton) 1971. Potential output and sources of economic growth: Korea, 1962–1968.

The study focuses on estimating potential output in Korea and identifies the sources of growth during the period 1962 to 1968. Three methods (Drabble, Adelman, and Denison) are employed in generating potential output of both sectoral and aggregate levels. The various growth rates of inputs and their contributions to economic growth are computed. Moreover, required investment to achieve full employment level of output are estimated. Finally, a comparison of various sectors is made to draw policy implications.

MICHAEL O. LERNER, Ph.D. California (Berkeley) 1970. Employment aspects of economic development in Brazil.

JOSEPH A. LICARI, Ph.D. Southern California 1970. On the dynamics of planned economy.

Growth and instability are two criteria upon which to base comparative evaluation of economic systems. This study was undertaken to better understand the character of these dynamic phenomena in a centrally planned economy. Review of the postwar historical record leads to a cyclical theory of planned economy growth based on concepts of supply-limitation and optimum plan tautness. General analysis of planned economy performance and detailed evaluation of the Hungarian experience yield little evidence to suggest that economic growth and instability are not joint products.

WILLIAM J. LOEHR, Ph.D. Colorado 1971. Toward a comprehensive evaluation of development institutions: the institutional effectiveness of the Nicaraguan rural credit program.

MILTON D. LOWER, Ph.D. Texas (Austin) 1970. Institutional bases of economic stagnation in Chile.

The dissertation uses the Veblen-Ayres theory of development to interpret Chilean stagnation in the 1950's. It furnishes an evolutionary accounting of the Chilean failure, despite a concerted industrialization effort after 1940 to achieve self-sustaining development. The introduction and diffusion of technical knowledge was insufficient to overcome severe institutional inhibitions to change. The institutional bases of stagnation are analyzed in terms both of historical survivals, such as the hacienda system, and of dysfunctional institutional adaptations of the system under the impact of borrowed technology and institutions.

JAMES L. MCCABE, Ph.D. Pennsylvania 1970. Two non-linear planning models of Turkey.

ERNESTO P. MARCOS, Ph.D. Notre Dame 1970. An investment planning model for Mexico.

This dissertation is concerned with developing a multisector, open-economy model to prescribe patterns of investment allocation consistent with several policy goals. Adopting the method of successive approximations in model building, a final version is designed to tackle specific policy problems confronting the Mexican economy. Under one formulation, a target rate of growth and a ceiling on foreign debt are given simultaneously. The findings suggest that manufacturing and agriculture should receive the highest priority.

JOSE N. MARIN X, D.B.A. Harvard 1971. Industrial finance in Nicaragua: A study of financial practice in an underdeveloped economy.

By applying a quantitative approach to the study of financial practices in the private sector of a single, underdeveloped economy, this thesis sets out to clarify certain developmental processes and illuminate targets for future investigation. On a pragmatic level, businessmen in Nicaragua can benefit from this systematic analysis of local corporate problems.

JAMES D. MIETUS, Ph.D. Fletcher School 1971. The use of social profitability analysis: an analysis of selected projects in the Dominican Republic.

BASSAM N: MITRI, Ph.D. Arizona 1971. The role and implications of the service sector in Lebanese economy.

Lebanese dependence on foreign trade is examined for its impact on economic stability, income distribution, and employment. Emphasis is placed on the special problems associated with the role played by the services sector in international transactions.

THORVALD MOE, Ph.D. Stanford 1970. Demographic developments and economic growth in Norway 1740–1940: an econometric study.

The basic analytical viewpoint underlying this study is that variations in demographic phenomena are caused primarily by changes in two classes of factors: economic conditions and demographic composition. The study is carried out with the help of econometric models in which variables describing economic conditions and demographic structure are assumed to be predetermined. The various models of economic-demographic interaction successfully predict the development of births and emigration in pre-World War II Norway.

PETER J. MOONEY, Ph.D. North Carolina 1970 The impact of immigration on the growth and development of the *U.S.* economy, 1890–1920.

This paper measures the levels and rates of change of several economic indicators which would have prevailed during the period 1890–1920, if the influence of the foreign stock was eliminated. A model is constructed to deal with the impact of labor force characteristics of the foreign stock on overall labor productivity—occupational preferences, sex composition, skill levels, and age distribution. The combined effect of these characteristics of the immigrant labor force was to increase the level of total labor productivity.

TAPAN MUKERJEE, Ph.D. Colorado 1970. Economic impact on Britain of decolonization of India.

WALID C. NIAZY, Ph.D. Pennsylvania 1971. The economics of public projects in less developed countries: a case study of the Aswan Dam.

INDER P. NIJHAWAN, Ph.D. North Carolina 1971. Asset preference and economic development with special reference to India.

This study tests the hypotheses: 1) there is preference for socially unproductive assets in underdeveloped countries in general and India in particular; 2) the predirection for these assets is inimical to growth potential of these countries. The evidence regarding portfolio selection points towards misdirection and misallocation of resources. Investment in these low priority assets adversely affects the growth rates of underdeveloped countries through effects on aggregate demand, saving, price level, growth and allocation of entrepreneurial skill, and organization of resources.

ROGER D. NORTON, Ph.D. Johns Hopkins 1971. A linear programming study of regional investment allocation.

AUGUSTINE M. NWABUZOR, D.B.A. Harvard 1971. Local foodstuff distribution in developing nations: a study in performance and opportunities in Nigerian wholesale trade.

Despite recent gains in food production, fear of Malthusian overpopulation and possible starvation in some developing nations persists. It has been observed that gains in food production—the Green Revolution—have engendered many problems, most of them marketing oriented. This has led to considerable research interest in the local food distribution systems of developing nations where a number of weaknesses have been observed. This study examined the extent to which these weaknesses exist in the Nigerian local food distribution systems.

CAROL C. OLIVEN, Ph.D. Iowa 1971. Some fiscal implications of regional integration: Latin America.

Traditionally, Latin American governments rely heavily upon customs receipts in their revenue structure. This study estimates the tariff-revenue adjustment resulting from intra-regional trade liberalization within the Latin American Free Trade Area. A generalized model designed to calculate the revenue impact associated with specific levels of tariff reduction is derived and modified to accommodate data available for eleven governments. The results indicate the pattern of revenue alterations which can be expected during entry.

MANSOUR OUBOUZAR, Ph.D. Kansas 1970. Educational planning and economic growth.

Educational planning has usually been treated as a residual problem in economic growth thus far. The purpose of this study was to analyze this problem in a more integrated way. Human capital formation in the form of education constitutes one sector among the other sectors of the growth model. Using Halanay's First Theorem on optimal control for systems with time lags I derived the necessary conditions for optimality in four cases.

JOHNSON A. OYELABI, Ph.D. Columbia 1970. Production functions and costs in Nigerian manufacturing industries.

This dissertation is an attempt to analyze quantitatively the production features of the Nigerian manufacturing sector. Concepts from the neoclassical theory of the firm are employed, with attention being focused on such aspects as economies of scale, the elasticities of factor substitution, output-employment relations and cost structure. 1965 data are used to estimate production functions of both the Cobb-Douglas and the *CES* varieties.

EDUARDO J. PADRON, Ph.D. Florida 1970. Economic development in Colombia since World War II.

GEORGE W. PARKER, Ph.D. Iowa 1971. Labor aspects of economic integration: Latin America.

This study uses three areas of economic theory: 1) labor mobility; 2) economic integration; and 3) social security. The theory was applied to the existing social security systems currently found in Latin America. Labor mobility depends upon the incidence of both the costs and the benefits of the particular social security programs in the various systems. These programs constitute potential trouble areas which may require harmonization in the event the potential becomes a reality.

JOHN L. PISCIOTTA, Ph.D. Texas (Austin) 1971. Development policy, inflation, and politics in Chile, 1938–1958: an essay in political economy.

This study examines the interplay of economic (development and anti-inflation) policies and politics. For each of the period's four presidents, consideration is given to his election campaign, efforts to form and maintain effective governing coalitions in the country's multiparty democracy, and the economic policy of his administration. The concluding chapter deals with implications of the Chilean case for selected theories of development and inflation, focusing attention on government decision-making capacity.

EDWARD J. RAY, Ph.D. Stanford 1971. Finance in a development context: lessons from West Germany.

Traditional explanations of slow growth in underdeveloped countries have led to inappropriate policy prescriptions. The view that taxes, subsidies or tariffs can promote rapid economic growth without creating gross inefficiencies is rejected. We hypothesize that fragmented capital markets have misallocated capital and repressed new investment. Our theory and empirical evidence from West Germany suggest that financial development can play a key role in promoting economic growth in poor and more advanced countries.

RICHARD REIDINGER, Ph.D. Duke 1971. Canal irrigation and institutions in North India: micro study and evaluation.

HENRY REMPEL, Ph.D. Wisconsin 1971. Labor migration into urban centers and urban unemployment in Kenya.

The central concern of this study is to explain why some men choose to migrate while others choose to remain in their present location and the effect these decisions have on urban unemployment. The study is restricted to in-migration into urban centers within Kenya and is based on both census data and origina data on migration flows and income levels obtained through the use of a survey.

EDWARD B. RICE, Ph.D. Massachusetts Institute of Technology 1971. Extension in the Andes.

GIUSEPPE C. RUGGERI, Ph.D. State University of New York (Albany) 1971. Automatic stabilizers and economic growth.

The Wharton Econometric Forecasting Model was used to simulate time-series of real output and capita stock in the United States from 1953 to 1963 first with and then without automatic stabilizers to determine the extent to which these stabilizers may act as a drag on economic growth. The results indicated the effects were negligible, except when mild recessions were followed by strong and sustained recovery periods.

JUSTINIAN F. RWEYEMAMU, Ph.D. Harvard 1971. An industrial strategy for Tanzania.

The dissertation denies that the grafting of capitalist institutions and the utilization of capitalist instruments in policy formulation and implementation (especially with respect of industrialization) will lead to self-sustaining and self-generated growth of the periphery. It asserts and verifies (using Tanzania as a case study) that due to their historical past, which is connected with the internationalization of capital, industrialization of these countries on the basis of the classical capitalist model produces "perverse" growth. Decentralized socialist industrialization is offered as a measure of overcoming underdevelopment.

VISHAL SABHERWAL, Ph.D. Oregon 1970. Planning in a labor surplus economy.

SAKR A. SAKR, Ph.D. Michigan State 1970. Development alternatives in Egypt in 1974–75.

A macro-economic model for Egypt that can be used for the purposes of projection planning is constructed so as to reflect the structural characteristics of the economy. The solution of the model makes it possible to generate some feasible and consistent sets of alternative development programs, each subjected to different values for the policy variables. The model could also be used for the identification of bottlenecks and alternative policy choice.

SHAPOUR SAMII, Ph.D. Lehigh 1971. Measurement o the predictive reliability of manufacturing plant and equipment expenditure models—an empirical approach.

MARK C. W. SCHROEDER, Ph.D. Cornell 1971. The impact of the Sonauli-Pokhara highway on the regional income and agricultural production of Pokhara Valley, Nepal.

The purpose of this study was to examine the impact of a recently completed highway on an isolated valley of Western Nepal. Estimates were made as to the role of transportation in the adoption of agricultural technology and the potential for increasing regional income.

JOHN A. SHAW, Ph.D. Purdue 1969. Commercialization in an agricultural economy: Fresno County, California, 1856–1900.

The evidence presented in this thesis is supportive of the ability of primary products to generate rapid and sustained economic growth during the nineteenth century in the mid-twentieth century's premier agricultural county. Primary data is presented for the capital market and estimates of gross income from agriculture, value added by manufacturing, income under alternative transportation systems, capital formation, and population changes are given. Changing markets and economic institutions are also discussed.

YUSUF SHIBL, Ph.D. California (Los Angeles) 1971. The Aswan High Dam: benefit and cost analysis.

Different investment criteria, implied by the economic theory of choice over time, are applied to the benefit and cost streams of the project to conclude whether it is a sound project or not from an economic efficiency point of view. The present value, the internal rate of return, and the ratio of discounted benefits to discounted costs are computed under various assumptions about the size of benefits and costs.

JOSÉ H. G. SILVA, Ph.D. Rice 1971. Dual economy and disguised unemployment.

The thesis develops a model for a subsistence economy. As a result, in a "latifundia" culture there is both declared and disguised unemployment. The thesis revises the notion of disguised unemployment by making use of the distinction between labor and labor services. In a labor surplus situation, with the workers' subsistence requirement, there is no general equilibrium and use is made of the notion of the "core" to show that the market solution is the only "stable" outcome.

JAMES L. STARKEY, Ph.D. Boston College 1971. Tax incentives and investment behavior in manufacturing, 1954–1968.

Since 1954 a number of tax incentives were offered to stimulate investment expenditures. This thesis is an attempt to determine the effectiveness of these incentives in the manufacturing industry. The analysis is conducted in the context of a simultaneous equation model of investment, finance, and dividend behavior. The impact of the incentives on actual expenditures is derived by estimating the stimulus to capital appropriations and then translating appropriations into expenditures. The results indicate that the incentive to expenditures was less than the revenue loss.

JOHN D. STULIGROSS, Ph.D. Oklahoma 1971. Planning by federal governmental agencies in the state of Missouri.

Planning has recently been adopted by federal, state, and local governments. This study considers federal planning within the state of Missouri. The principal source of data is private interviews with federal executives in the state. Some federal departments in Missouri depend on plans prepared by state agencies; some plan internally; others do not plan. The departments which plan recognize the value of this function and expect to utilize it even more in the future.

MOISES SYRQUIN, Ph.D. Harvard 1971. Production functions and regional efficiency in the manufacturing sector in Mexico, 1965.

This study estimates aggregate production functions for the manufacturing sector of Mexico from cross section data, at different levels of geographic and industrial aggregation. Cobb-Douglas and *CES* type functions are estimated for total manufacturing across counties at different levels of geographic comprehensiveness, and for two- and three-digit level sectors across states. Regional differentials in efficiency are analyzed and the interstate variation of an efficiency index derived from covariance analysis, is analyzed by regression and factor analyses using over thirty regional variables.

LESLIE SZEPLAKI, Ph.D. California (Berkeley) 1970. Hungarian economic planning before and after the 1968 reforms, with an emphasis on financial institutions.

OM P. TANGRI, Ph.D. California (Berkeley) 1971. India's program of agricultural development 1947–67: an analysis and appraisal.

This study examines the major problems that have impeded economic development in India's agricultural sector between 1947–67. Such data as were available for the period 1967–69, were also used where appropriate. From the standpoint of policy, the purpose of the study lies more in suggesting future directions along which changes are needed in Indian agricultural planning, than in prescribing a multitude of specific, tailor-made programs.

IZUMI TANIGUCHI, Ph.D. Texas (Austin) 1970. The financing of Japanese imports with especial reference to foreign investments in the process, 1868–1914.

The importation of Western technology, raw materials and consumer goods was basic to Japan's economic development. The transfer of technology was effected mainly by bringing Western technicians to Japan as instructors and by sending their own students to study abroad. Essential imports were paid for, mainly, by increasing the production and exports of raw silk. In the beginning specie export was also important. Later the export of cotton textiles and shipping services became increasingly important. The role of foreign investments was minor.

Firouz Vakil, Ph.D. California (Berkeley) 1970. The role of the distribution of income in economic development: some empirical evidence.

Douglas O. Walker, Ph.D. Southern California 1970. Socialist price formation models and input-output analysis: a study of Hungary, 1959–1964.

In this study a number of price-formation models that have been discussed in eastern Europe are presented and applied to the Hungarian economy. The study focuses on contrasts between the view of the economy reflected by the original price system and that created when flows of economic activity are revalued to conform more closely to costs associated with production. The comparison is made mainly on the basis of input-output tables.

Morris D. Whitaker, Ph.D. Purdue 1970. Labor absorption in Brazil: an analysis of the industrial sector.

The objective of this study was to examine the Brazilian labor market in the postwar period and to explain the marked decline in labor absorption in Brazilian industry. Econometric models of the labor market serving the manufacturing sector were estimated. Results suggested that 80 percent of the decline in industrial sector capacity to absorb labor was due to the traditional sector, contrary to popular opinion. Econometric results were used to evaluate alternative labor market policies.

Ann D. Witte, Ph.D. North Carolina State 1971. Employment in the manufacturing sector of developing economies: a study of Mexico, Peru, and Venezuela.

An investigation of the hypothesis that the relatively slow growth of employment in the manufacturing sector of developing economies is largely the result of production functions with relatively high elasticities of substitution and a shift in relative prices which makes it profitable for firms to substitute capital for labor. This hypothesis was substantiated for all industries studied in Mexico and the majority of industries studied in Peru.

Hossein Yaghoobi, Ph.D. Michigan State 1971. The most efficient production techniques for providing nutrition and income from the agricultural sector of Nigeria.

This dissertation considers the twin objectives of earning income from the cash crops and obtaining nutrients from the food crops. It applies a mathematical programming model to the agricultural sector of Nigeria. To find answers for several questions important to policy makers, the model examines the various production techniques of crops in different stages. It first introduces improved practices of crop production. Secondly the model examines prospective varieties of crops. Thirdly, it imposes maximum limits on the supply of fertilizers and extension services.

Ergun Yener, Ph.D. North Carolina 1970. Financial and capital market development in Turkey: problems and prospects.

Suk Bum Yoon, Ph.D. Pennsylvania 1971. A macroeconometric analysis of the impact of foreign capital investment in an underdeveloped country: the case of Korea.

An intermediate size econometric model (twenty-seven stochastic and twenty identity equations) is established for the Korean economy 1) to describe the structure of the economy, 2) to predict the possibility of "self-sufficiency" of the economy, and 3) to analyze the impact of foreign capital investment. *OLS* and psuedo-*SOIV* estimators are employed in parameter estimations, and a Gauss-Seidel technique is used in simulations. The results are obtained as expected.

Vernon W. Yorgason, Ph.D. Kansas State 1971. Agriculture in Canadian economic development: an econometric approach.

Mathematical and econometric techniques are utilized to explain the effects of government expenditure on agricultural research and extension on agricultural output, and the resulting change in agricultural output on gross domestic product at factor cost. A thirty equation long-run agriculture-industry macroeconomic model is developed to quantify the pertinent relationships. The North linkage hypothesis is used to explain intersectoral shifts in this system. Agricultural development is apparently a highly significant factor in Canadian economic development.

Habib A. Zuberi, Ph.D. Southern Illinois 1971. Institutional financing and balanced growth: an empirical case study of Pakistan.

This study provides systematic analysis of growth of financial institutions in Pakistan and establishes the relationship between institutional credit and output in the industrial and agricultural sectors. A close relationship was found between the prices of food and manufacturing goods. The ratio of institutional credit was strongly correlated with price ratios and output ratios in the respective sectors. The government, by controlling institutional credit, maintained investment incentives, stable price ratios, hence balanced growth.

Economic Statistics; including Econometric Methods, Economic and Social Accounting

Amotz Amiad, Ph.D. California (Berkeley) 1970. Theoretical and empirical aspects of the economics of information in the sorting of industrial raw materials.

Terry N. Barr, Ph.D. Washington State 1971. The constant elasticity of substitution production function and its applicability in applied research.

The objective of this thesis is to demonstrate the theoretical superiority of the *CES* function and then carefully develop a technique for reducing its statistical complexity. The technique is utilized in an empirical situation in an attempt to obtain suggestive evidence as to the applicability of the *CES* production function in applied research under the technique.

RICHARD E. BENNETT, Ph.D. Ohio State 1970. An information system for economic analysis.

The problem is to develop a system which can be used to identify changes in the state of the economic environment. To attack this problem a series of social accounting matrices is developed which sets forth a consistent set of aggregate economic data. The social accounting matrices provide the basis for determining the essential features of whether or not the state of the economy is maintained or altered by the use of principal components.

ERIC C. BLANKMEYER, Ph.D. Princeton 1971. Dynamic portfolio analysis and export stabilization.

Static portfolio analysis is found to be inadequate when the risky assets have a lagged auto- or cross-covariance structure. When a quadratic objective function is maximized over several time periods, this is equivalent to filtering the asset prices through Bartlett's spectral window which accentuates long-swing variance and attenuates short swings. An application is made to the stabilization of a portfolio of commodity exports, and opportunities for diversification are given a frequency-domain interpretation.

F. CHARLTON BRAITHWAITE, Ph.D. Queen's 1971. An econometric analysis of the determinance of investment in Canadian manufacturing.

This thesis analyzes the determinance of growth investment in Canadian manufacturing during the years 1946–64 using a modified neoclassical investment model. Annual data are used to estimate the model for twelve manufacturing industries and total manufacturing by the method of ordinary least squares and for total manufacturing by the method of structurally ordered capital instrumental variable developed by Franklin Fisher. Use is made of input-output relationships in the application of the latter method of estimation.

JOHN P. BROWN, Ph.D. California (Berkeley) 1970. The economic effects of floods: investigations of a stochastic model of rational investment behavior in the face of floods.

CAROL S. CARSON, Ph.D. George Washington 1971. The history of the *U.S.* income and product accounts: the development of an analytical tool.

Influential factors were investigated—events, issues, theories, individuals—in the development of national income from a measure of production into income and product accounts, seen as a body of interrelated aggregates designed to facilitate analysis of the functioning of the economy. "Oral history" was extensively utilized for the post-1930 period. Uses, and also sources and methods, were documented at several points for perspective on future directions the accounts may take within broader schemes of economic and social accounting.

ROBIN A. L. CARTER, Ph.D. Queen's 1971. Small sample properties of three econometric methods of estimation.

Monte Carlo experiments examined the small properties of: Ordinary Least Squares (*OLS*), Two-Stage Least Squares (*2SLS*), Simultaneous Least Squares (*SLS*) and a new method of estimating their standard errors. The ranking of the estimators on the criteria of bias and root mean square error varied with the model used. Either *SLS* or *2SLS* ranked first with *OLS* third. *SLS* always had the best standard error estimates and the most accurate forecasts.

C. SCOTT CLARK, Ph.D. California (Berkeley) 1971. The short-run demand for employment in durable goods industries.

KENNETH O. COGGER, Ph.D. Michigan 1971. Statistical foundations of general order exponential smoothing with implications for applied time-series analysis.

NEIL R. CONN, Ph.D. Duke 1971. The construction, fitting, and application of homogeneous production functions in the estimation of disembodied technical progress.

PRUDENCE S. CROZIER, Ph.D. Harvard 1971. Price equations at the retail level.

RAYMOND DANIEL, Ph.D. Purdue 1970. An econometric analysis of the farmer demand for fertilizer nutrients.

The primary objective of this study was to identify major factors which affect quantity of fertilizer demanded seasonally and to estimate demand relationships for individual farmers. Data used in this study represented fertilizer purchases and demographic characteristics reported by a representative panel of approximately 900 Illinois farmers (1961–65). Three different fertilizer demand models were estimated for spring and fall: 1) total fertilizer; 2) all fertilizer nutrients; and 3) straight-mixed fertilizer nutrient models. The last model appeared most appropriate and consisted of four independent equations (straight N, P_2O_5, K_2O, and a mixed fertilizer equation).

JOHN A. DUFFY, JR., Ph.D. Boston College 1971. The influence of the income-recipient unit on the measurement of inequality in the distribution of world income.

This influence is estimated by constructing and comparing three distributions of per-unit national incomes, using the U.N. national income concept in conjunction with the individual person, the household, and the "standard consumer unit" as income-recipient units. Employed are 1958 data for ninety-one countries, exchange-rate conversion, and the Lorenz-curve concentration ratio approach for summarizing inequality. Recipient-unit variation is found to influence world income inequality estimates by as much as 12.5 percent in relative terms.

Donald H. Ebbeler, Ph.D. Purdue 1970. An investigation of the properties of exact finite sample distributions of *GCL* statistics associated with the structural representation of an econometric model.

Some properties of the distribution functions of the *GCL* identifiability test statistic, endogenous variable structural variance estimators for a class of simultaneous equations explanatory econometric models are demonstrated. Asymptotic expansions are employed to approximate the finite sample distribution functions. Some new results on *k*th order confluent hypergeometric functions are presented.

Duane O. Eberhardt, Ph.D. Southern California 1970. An econometric overhead model of an aerospace firm.

This study consists of the development and validation of an econometric model for forecasting overhead accounts in an aerospace firm. The model was validated by comparing its forecasts with actual data and also with forecasts made by the subjective method currently being used by the firm's accounting department, and with naive forecasts. It was found that the model generally was able to predict historical and future values better than the naive models, and could serve as an important forecasting aid for the accounting department.

Mary L. S. Epps, Ph.D. Duke 1970. A computer simulation of the world coffee economy.

Melvin A. Fuss, Ph.D. California (Berkeley) 1970. The time structure of technology: an empirical analysis of the putty-clay hypothesis.

Charles M. Hamilton, Ph.D. Kansas 1970. Specification error in the simultaneous equations model.

This paper investigates specification error caused by the omission of exogenous and endogenous variables and its consequent effect on the exact finite sample distribution properties (bias, mean square error) of various estimators in the simultaneous equations model. The parameters are estimated by both ordinary least squares and two-stage least squares which permits, in addition to the above, an analysis of the relative sensitivity or robustness of the two estimation procedures.

Michael J. Hartley, Ph.D. Pennsylvania 1970. Estimation and prediction in autoregressive models: a new approach.

A general class of so-called "minimum distance" estimators is explored for single equation as well as simultaneous equation autoregressive models. In the former situation, least squares is shown to be a member of the class. Another member, the final form estimator, is shown to optimize the dynamic simulation path of the endogenous variables over the sample range. Other members of the class are within-sample analogues of prediction problems beyond the range of the sample.

Medhat A. Hassanein, Ph.D. Pennsylvania 1970. Towards an econometric model of some sectors of the *UAR* economy, 1946–1966.

The model consists of 146 equations; 80 stochastic functions and 66 identities. The *UAR* economy was divided into five principal sectors. Each sector was studied in terms of disaggregated commodity and factor markets. New series for components of aggregate demand and supply were constructed. Several statistical techniques were applied. A neo-Keynesian macroeconomic type seems to fit the Egyptian experience, in spite of the comprehensive central planning in the 1960's.

John W. Helmuth, Ph.D. Missouri 1970. Futures trading under conditions of uncertainity.

The goal of this study was to examine the operation of futures markets within the framework of the economic behavior under conditions of uncertainty. Theories of expected utility maximization and risk aversion were used as foundations for characterizing futures trading in a portfolio framework. Analysis of variance was used to systematically partition futures price variance into relevant components and regression techniques were then used to explain the components of variance.

Dun-Mow Hong, Ph.D. Ohio State 1971. The effects of autocorrelated errors on several least squares estimators.

The purpose of this dissertation is to shed light on the small sample properties of various least squares estimators of the parameters of several types of linear regression models having autocorrelated errors. The approach consisted of conducting sampling experiments. A first-order Markov process was assumed for the autocorrelated errors.

Keith H. Johnson, Ph.D. Illinois (Urbana) 1970. On combining time-series and cross section data in dynamic economic models.

Robert R. Keller, Ph.D. Wisconsin (Madison) 1971. An analysis of relative income shares in the United States during the 1920's.

Relative factor income shares were measured for 1923 and 1929 and compared to other prosperity periods in the twentieth century. The 1920's was the only decade where capital's share of income increased. The basic explanation for the increase in capital's share in the metals, railroad, and electric utility subsectors was that the price of output tended to be inflexible in the presence of falling costs of production. Although there were other sources of cost reduction, the role of declining fuel costs was emphasized.

James B. McDonald, Ph.D. Purdue 1970. An analysis of the properties of the finite sample distribution of some structural parameter estimators.

Robert F. McNown, Ph.D. California (San Diego) 1971. A simulation study of the small sample properties of several distributed lag estimators.

Four techniques for estimating distributed lag patterns of general shapes were analyzed by Monte Carlo methods. The four techniques were (i) the rational lag—ordinary least squares, (ii) the rational lag—weighted least squares, (iii) the polynomial lag, and (iv) stepwise ordinary least squares. The techniques were applied to eight different patterns ranging from those with constantly decreasing weights to some with several turning points, and their relative performances were studied under seven alternative environments involving combinations of such statistical complications as multicollinearity, autocorrelation and limited degrees of freedom.

ONO MITSUO, Ph.D. George Washington 1971. The marketing accounts system: an introductory analysis.

This study formulates an economic accounting framework, designated the marketing accounts, which is used to trace systematically the flow of sales transactions completed among various types of retailers and of households (excluding nonprofit institutions and trust funds). An empirical implementation of the marketing accounts suggests that, currently, it is useful not so much as an analytical device but rather as a framework to be used to reconcile statistical data.

CARL C. MOORE, Ph.D. Alabama 1970. An evaluation of applications of a mathematical simulation model in applied socioeconomic research.

LOUISE NORDSTROM, Ph.D. Clark 1971. Economic forecasts and information theory.

Using information theory, the accuracy of nineteen yearly economic forecasters of *GNP* and its components are evaluated on two bases: distribution and turning points. The results are compared to accuracy in predicting the level of *GNP*. The information measure of distribution accuracy does not correlate with level accuracy and indicates that forecasters have improved over time. Furthermore, turning point increases are predicted better than decreases and *GNP* turning points better than component turning points.

MAKOTO OHTA, Ph.D. California (Berkeley) 1971. Hedonic price index for boiler and turbo-generator: a cost function approach.

GEORGE N. PARASKEVOPOULOS, Ph.D. Pennsylvania 1971. An econometric analysis of the foreign trade of Greece.

Utilizing annual and quarterly data (1954–1966), Greece's import and export demand equations for major groups and subgroups of commodities and services have been estimated. The study revealed that Greece's imports are income and price elastic, commodity exports to *OECD* countries are income inelastic, most exported commodities face strong price competition, and foreign demand for tourist services is highly income and price elastic. Applications of the estimates have been made, and policy implications have been analyzed.

RAYMOND A. PEPIN, Ph.D. Clark 1971. A Monte Carlo study of the small sample properties of several simultaneous equation estimating techniques when statistical problems are present.

A number of sampling experiments are used in studying the small sample properties of three simultaneous regression equation estimating techniques: ordinary least squares, two-stage least squares, and limited-information single equation. The results suggest that no method is clearly superior to the others when several statistical problems are present. In addition, the usual procedure followed by economic model builders, which relies on the use of ordinary least squares estimates in specifying the model's structural relationships, is shown to be undesirable.

A. A. ILMARI PIETARINEN, Ph.D. Duke 1971. On the measurability of structure in economics.

RICHARD D. PORTER, Ph.D. Wisconsin (Madison) 1971. Strategies for discrete-time decision models.

We explore some discrete-time decision models with general continuous outcome and payoff functions, and develop and analyze some new methods for computing solutions to such models.

AKUNDI S. RAO, Ph.D. Rochester 1971. Likelihood methods for the estimation of distributed lag models.

Methods are developed for the direct maximum likelihood estimation of Solow's and Jorgenson's distributed lag models. These methods are then compared with the estimation of the two models in their autoregressive forms, as is usually done. A Bayesian analysis of the Solow model is discussed. The power function of Durbin's recent test, for serial correlation in regressions with lagged dependent variables, is studied and compared with that of the likelihood ratio test.

JAMMULA M. REDDY, Ph.D. Alberta 1970. Statistical estimation of the elasticities of demand for nondurable goods in the Canadian economy.

JEFFREY A. ROTH, Ph.D. Michigan State 1971. An econometric approach to technological change and returns to scale in steam-electric generation.

The conclusions of this thesis are that thermal electric plants operate under increasing returns to scale, that technological change between 1948 and 1965 affected plant efficiency only by permitting construction of larger plants, and that the ratio of plant capacity to output is affected by variations in the number of turbogenerators per plant. Supporting evidence results from estimation of a fixed-relative-proportions model with exogenous output previously tested and found free of specification error.

DANIEL C. SCHEEL, Ph.D. Oregon 1970. Testing for autocorrelation in rational distributed lag models.

PETER SCHMIDT, Ph.D. Michigan State 1970. Regression analysis with second-order autoregressive disturbances.

The testing and estimation procedures which are used to alleviate the problems caused by autocorrelated disturbances usually assume that the disturbances follow a first-order autoregressive process. This study attempts to develop suitable testing and estimation procedures under the assumption of second-order autoregressive disturbances. A number of results (obtained mainly by Monte Carlo methods) suggest that this added generality may be useful unless one has a priori information which indicates a first-order process.

Venkataraman Sundararajan, Ph.D. Harvard 1971. A flexible coefficient bridge model: trend-cycle adjustments in input-output analysis.

Dalip S. Swamy, Ph.D. Pennsylvania 1970. An econometric model of the *U.S.* financial markets.

Using fifty-four structural equations and a number of identities, this model evaluates the effects of changes in certain monetary policy variables. A change in the *Q* ceiling rate is found to alter the allocation of credit between banks and nonbank financial institutions. Restricting the use of Euro-dollar borrowings leads to unloading of Government securities by banks and a possible twist in the interest rate structure. A change in unborrowed reserves produces sharp effects on short-term interest rates.

Peter A. Victor, Ph.D. British Columbia 1971. Input-output analysis and the study of economic and environmental interactions.

The theoretical core of the thesis uses input-output analysis to study the flow of materials between an economy and the natural environment. Two basic models are developed. Both employ a commodity-by-industry classification. Data on the waste products and 'free' goods produced and used by the Canadian economy in 1961 are estimated and collected. This data is combined with input-output data and some estimates are made of the environmental impact of Canadian economic activity.

Frenck Waage, Ph.D. California (Berkeley) 1971. Transient and steady state decision rules of a Bayesian consumer who buys in binomial markets.

Michael A. Walker, Ph.D. Western Ontario 1971. A long-term econometric model of the Canadian economy with emphasis on the financial sector: a simulation study of some alternative policy rules.

A non-linear model consisting of forty-six simultaneous equations is estimated from an annual sample from the postwar period using limited information techniques. Extensive simulation and reestimation is carried out to obtain a 'best' dynamic structure. A control sequence and single period dynamic forecast beyond the sample is presented. Money-supply and interest-rate rules are simulated and ranked according to their effect on fifteen-year growth profiles. Ranking was not affected by random shocks.

David A. Wyss, Ph.D. Harvard 1971. An econometric model of prices and wages in *U.S.* manufacturing.

This thesis presents a model of wages and prices at the two-digit SIC level. In the wage model, the interactions between industries are the major force on wage changes. This creates a model with significantly longer lags than more aggregate forms. For the price model indices of input and output prices were constructed fo each two-digit industry. The industries were the separated into behavioral categories agreeing with thei concentration.

Arshad Zaman, Ph.D. Michigan State 1970. Consume demand in the United States 1929–65: analysis an estimation.

An investigation of the principles underlying th formulation and estimation of a complete system o demand equations. Analysis of existing functional form and two suggestions for theoretically plausible parametrizations of the double-logarithmic specification fo demand equations. These include a *CES* specification Use of modified Kmenta-Gilbert procedure for obtaining restricted Zellner-Aitken estimates of parameter from *U.S.* data on personal consumption expenditures Estimates compare favorably with those obtaine under leading alternative specifications.

Monetary and Fiscal Theory, Policy, and Institutions

Ruben D. Almonacid, Ph.D. Chicago 1971. Nomina income, output, and prices in the short run.

A theory is developed concerning the short-run behavior of output and prices in which the deviations o the economy from equilibrium output are explained b the failures of both laborers and entrepreneurs to hav accurate anticipations and/or perceptions about prices The aggregate supply curve is shown to be positivel sloped when laborers misjudge consumer prices an entrepreneurs misjudge input prices, but perfectl inelastic when all prices are judged accurately. Th results of tests using *U.S.* data are favorable.

Elias R. Ason, Ph.D. California (Berkeley) 1970. A econometric model of the Puerto Rican bankin sector.

Randall Bartlett, Ph.D. Stanford 1971. A positiv view of government actions in a market economy.

The dissertation examines the rational response of vote maximizing government to the pressures place upon it by self-interested agents in an uncertain world It provides a theortical explanation for the observe patterns of purchasing and taxing generated by democratically organized political system and concludes that public actions in the market, like the allocation of resources in the private sector, will reflect th distribution of wealth.

William S. Becker, Ph.D. Louisiana State 197 Agricultural credit and Columbia's economic development.

The purpose of this research is to examine the theoretical and practical cases for focusing developmenta efforts on Colombia's rural sector as a source of ec nomic expansion. Improvements in agricultural pr ductivity, it is shown, can have beneficial effects on th process of capital formation, the balance of payment and the development of an industrial base. It is argued that the necessary productivity improvements in the agricultural sector can readily be accomplished by modifications in the provision of agricultural credit.

H. PRESCOTT BEIGHLEY, Ph.D. Purdue 1970. An application of linear programming to commercial bank asset management under certainty and uncertainty.

A general multiperiod linear programming profit-planning model of commercial bank asset management is constructed and applied to asset management problems under conditions of certainty and uncertainty. Hypothetical data are derived and used to obtain numerical solutions, and policy questions of interest to the banker are discussed.

ROSALIND BLAUER, Ph.D. McGill 1971. Inflation and the distribution of income and net worth of Canadian households: 1950–1967.

JAMES D. BLUMENTHAL, Ph.D. Columbia 1970. Expectations, forward rates, and the term structure of interest rates: municipal bonds.

WILLIAM E. BOYD, Ph.D. Florida State 1971. Response to monetary policy: 1960–1969.

This study tests the hypothesis that lags in the effect of monetary policy vary between Federal Reserve Districts owing to differences in financial characteristics of commercial banks. The response of selected monetary variables to a change in monetary policy was found to be highly variable within each district, but little significant difference was found in the average lag for each of the twelve districts.

FRANCIS F. BREEN, JR., Ph.D. Boston College 1971. The demand for currency: a multinational study.

Postwar data for twenty-four countries are employed in testing hypotheses about the demand for currency suggested by a cash-inventory model and by previous studies of the demand for money and for currency. Individual-country and pooled-data regression estimates show unitary income elasticity predominating, with some support for "economies of scale." Absence on money illusion seems quite general, and there is evidence of interest-sensitivity of currency balances in economically advanced countries.

IRA J. BROUS, Ph.D. Syracuse 1970. Towards the conversion of the public debt into an undifferentiated account at a uniform rate of interest.

EDGAR K. BROWNING, Ph.D. Princeton 1971. Income redistribution and the negative income tax: a theoretical analysis.

This dissertation is a theoretical analysis of income transfer programs with an emphasis on the negative income tax. Effects on work effort, distribution, efficiency, and collective decision making are stressed. The final chapter compares the negative income tax to a specific form of wage subsidy and the analysis suggests that this wage subsidy may be the more efficient method of redistribution.

JAMES B. BURNHAM, Ph.D. Washington (St. Louis) 1970. Household demand for financial assets, 1947–1967.

The dissertation investigates the impact of complementary asset-holding patterns on financial system stability. Using the theoretical guidelines developed by Brainard and Tobin, flow of funds data for the period of 1947–1967 and various interest rate series are examined with linear regression techniques to test for complementarity in the household sector. The probable existence of complementarity is shown and its implications are discussed.

DONALD R. BYRNE, Ph.D. Notre Dame 1971. An empirically based theoretical analysis of the typical savings and loan association.

Using cross-sectional data for the years 1960 through 1963 and multivariate regression analysis, the SMSA Detroit is studied to ascertain the crucial determinants of operating income and expense ratios. Primary emphasis concerns the operating expense ratio models in which average deposit size and *FHA* plus *VA* mortgage ratio variables are found statistically significant. The asset size variable does not generally yield significant t-values. Location seems a more important qualitative variable than does charter type.

GEORGE N. CARLSON, Ph.D. Illinois (Urbana) 1971. The effects of retail sales taxation of producers' goods on industry relative prices and values added.

This dissertation utilizes an input-output model of the *U.S.* economy to measure the extent of retail sales taxation of producers goods and estimate the effects of this taxation on industry relative prices and values added. The price effects depend on an industry's direct tax cost and the taxes embodied in purchased inputs. If firms are unable to shift tax costs, value added falls by more than the unshifted portion of the tax cost.

JON T. CAULEY, Ph.D. Colorado 1971. The short-run shifting of the corporate income tax: an empirical test.

CAROLYN CLARK, Ph.D. California (Berkeley) 1970. The permanent income hypothesis of the demand for money in Canada, 1926–1965.

TERRENCE M. CLAURETIE, Ph.D. Washington State 1971. Monetary growth rates, the business demand for funds, and the residential mortgage market.

SAMMIE K. COOPER, Ph.D. Texas (Austin) 1971. The economic effects of alternative uses of inactive state funds: the case of Texas.

A theoretical exploration of optimality in cash management for governmental units and an empirical analysis of the administration of state funds in Texas.

JAMES C. COX, Ph.D. Harvard 1971. Liquidity preference and the welfare economics of taxation and government expenditure.

Maureen S. Crandall, Ph.D. Northwestern 1969. Monetary instrument variables as regressors in expenditure functions.

The study seeks to establish the nature of the direct correspondence between the parameters of action of the monetary authorities and aggregate expenditure from 1947–64. Through several distributed lag patterns, *GNP* is regressed on such variables as reserve requirements, discount rate, Federal Reserve open market holdings, and Treasury gold holdings. The results suggest a monetary lag of a year or less, and that the strongest variables are those which work upon *GNP* through bank reserves.

Michael R. Darby, Ph.D. Chicago 1970. The dynamics of the allocation of transitory income among consumers' assets.

This dissertation presents an all-capital model taking account of the distinction between permanent and transitory flows and stocks. The model provides a transmission mechanism for monetary policy which is characterized by variable lags. The empirical results also confirm the implication that the marginal propensity to spend out of transitory income is variable and near its minimum at the cyclical peak and trough.

Luc De Wulf, Ph.D. Clark 1971. Taxation of unrealized capital gains: theoretical and empirical analysis.

Capital gains have long been given preferential treatment not only in the United States, but in Canada, the United Kindgom, Sweden, and Belgium as well. Reasons for this favorable treatment are briefly explored. Since such a favorable treatment of capital gains violates the principles of equity, alternative methods of capital gains taxation are proposed. The resulting tax revenues and effective tax rates by income classes are calculated (using 1966 data) and compared with the 1966 treatment.

Warren R. Dixon, Ph.D. Colorado 1971. The secular and cyclical income elasticity of the property tax base.

Steven Dobson, Ph.D. California (Berkeley) 1970. The term structure of interest rates and the government securities market.

Robert D. Ebel, Ph.D. Purdue 1971. Value added taxation in the subnational economy: an empirical analysis and evaluation.

The study examines in detail the background, statutory provisions, and structural and economic characteristics of the Michigan Business Activities (value added) Tax (1953–1968). The tax is examined for neutrality among factor inputs, impact on various business classifications, comprehensiveness of coverage, responsiveness to economic growth, certainty of shifting, ease of administration, and its effects on industrial location and growth. In addition, the revenue collection and interindustry tax liability effects of alternative formulas for apportioning multistate business receipts is empirically measured.

Robert A. Eisenbeis, Ph.D. Wisconsin (Madison) 1971. A study of the delineation of geographic markets for business loans.

In this study cluster and discriminant analysis techniques are used to delineate the geographic markets for business loans around eight major *U.S.* cities. Using data from the 1955 Federal Reserve Business Loan Survey an attempt is made to determine: 1) the number of distinct markets existing about each city; 2) the geographic extent of each market; 3) the characteristics of both the transactions and borrowing firms in each market; and 4) the relative economic significance of the markets for each city.

Robert H. Floyd, Ph.D. Rice 1971. A theoretical analysis of domestic tax systems and the provisions of the General Agreement on Tariffs and Trade for their application to international trade.

The thesis examines the implications for economic efficiency and tax neutrality of the provisions of the General Agreement on Tariffs and Trade which codify the international treatment of internal taxes on traded goods. In addition, several recently proposed changes in these provisions are examined. The study develops and employs a neoclassical, two-country, general equilibrium model for a theoretical analysis of the effects of various product and factor tax schemes on consumption, production, and trade patterns.

James S. Fralick, Ph.D. Syracuse 1970. A microeconomic analysis of the impact of changes in the depreciation—tax credit laws in the investment spending of twenty-six firms.

Benjamin M. Friedman, Ph.D. Harvard 1971. Methods in optimization for economic stabilization policy.

The primary objective of this thesis is to develop a methodology for applying econometric models to problems in government policy for economic stabilization. An extended dynamic optimization framework generalizes a Tinbergen-Theil algorithm to include a piecewise quadratic objective, inequality constraints, and an endogenous time horizon. The thesis also develops the concept of subinterval linearization of econometric models and derives such linearizations for the Wharton models. A final section uses the optimization algorithm and the linearized Wharton model to perform optimal stabilization exercises.

John G. Fulmer, Jr., Ph.D. Alabama 1970. An investigation of the effects of Federal Reserve system membership on individual commercial banks.

Louis M. A. R. Gevers, Ph.D. Stanford 1970. Money and growth under uncertainty.

This paper discusses the welfare implications of monetary policy in a general equilibrium neoclassical growth framework. Individuals face uncertainty in their own production processes and therefore have a portfolio demand for money as a safe asset. If individuals cannot trade risks, Pareto-efficient allocations may still be approximated by government action—suitable tax-subsidy policies must be established and the rate of deflation of fiat money must be close to the average return on risky assets.

ROBERT A. GILBERT, Ph.D. Texas A&M 1971. Deposit relationships and bank portfolio selection.

A theoretical model of bank portfolio selection is developed in which a bank offers lines of credit to some customers which are multiples of their deposits. The bank loses desposits if it does not meet the loan demands of these customers, which are random variables. The model is used to analyze the allocation of bank loans between depositors and nondepositors by an expected profit maximizing banker as monetary policy becomes more restrictive.

CAROL S. GREENWALD, Ph.D. Columbia 1971. Recession as a policy instrument: Israel, 1965–1969.

This dissertation examines the effectiveness of a recession in Israel during the mid-1960's as a policy instrument in restructuring production toward exports. Attention is focused on the Israeli trade balance, the role played by fiscal and monetary policies and on the domestic adjustments which facilitated or hindered the attainment of external balance. It was concluded that the recession did induce structural changes in the economy which allowed Israel to renew rapid growth with an industrial structure more aligned to its development needs.

DANESH C. GUPTA, Ph.D. Minnesota 1971. Monetary influences on the inflationary trend in India during the third five-year plan (1961–66).

The thesis is a quantitative study of the demand function for money in India. A linear function is estimated by using quarterly data for 1948–68. The entire period as well as five subperiods are analyzed to test the significance of explanatory variables. The findings for the most part support the existing theoretical formulations.

GIRDHARILAL S. GUPTA, Ph.D. Johns Hopkins 1970. A monetary policy model for India, 1948–49 to 1967–68.

Based on a sample of twenty annual observations (1948–49 through 1967–68), a policy-oriented econometric model for the Indian economy with an emphasis on the monetary sector has been formulated, estimated, and analyzed. The primary objective was to evaluate quantitatively the direct and indirect impacts of various policy variables on the model's endogenous variables. The model should be of help in understanding the portfolio management by the different sectors of the economy.

DONALD W. HOLLEY, Ph.D. California (Riverside) 1971. A two-sector neoclassical monetary growth model.

This is an attempt to set up a two-sector model (households and producers) and to empirically test whether changes in the ownership of money between the two sectors has any effect on the neutrality of money. Included is a discussion on the definition of money which concludes money is a medium of exchange and this characteristic is shared by many goods. The concept is quantified and used as the measure of money in the empirical work.

KENNETH W. HOLLMAN, Ph.D. Alabama 1970. An analysis of the Alabama state debt, 1955–1965.

DENNIS K. HOOVER, Ph.D. Columbia 1970. Problems in the implementation of economic stabilization policy, United States, 1965–1968.

The study utilized the inside time lag and investigated the short-range economic goals and instruments of major *U.S.* policy makers, communication of information, and coordination of policy. It was concluded that an excess of goals over the number of available, acceptable instruments and failures in communication and coordination contributed to the delayed fiscal and monetary actions of the period. Informational sources included interviews, Congressional hearings, and various government publications.

LOUIS J. JAMES, Ph.D. Kansas 1970. A general equilibrium analysis of the expenditure effects of federal grant-in-aid programs.

A two-state static general equilibrium trade model consisting of seven independent equations is totally differentiated with respect to an autonomous change in both matching and unconditional grants. Stimulation, substitution, and neutral effects are obtained under quantitative restrictions indicating the constant price micro analysis in the literature requires implicit assumptions unsuitable for a full employment economy, generally even when tax effects are considered. The conclusions also indicate changes in relative prices are generally mandatory.

IVAN C. JOHNSON, Ph.D. Western Ontario 1971. Provincial and municipal debt in Canada, 1946–66.

This thesis investigates: 1) the sensitivity of debt issue by the provinces and municipalities of Canada to federal monetary policy; and 2) the effect that the presence of provincial and municipal debt in the portfolios of financial intermediaries has on the efficacy of federal monetary policy.

FRANK J. JONES, Ph.D. Stanford 1971. Household acquisition of financial assets: an empirical investigation.

DUK CHONG KIM, Ph.D. Missouri 1970. The achievement of a government output objective through the use of subsidies.

An analysis of how a government may induce a level of output which is greater than a market rate of output through the employment of different methods of subsidy in short-run competitive and monopolistic markets. It finds that different methods result in expenditure of different amounts. The effects of the different subsidies on the efficiency of firms are also considered.

Peter G. Kirkham, Ph.D. Princeton 1970. The portfolio behavior of selected Canadian financial intermediaries: an econometric analysis.

An attempt is made to construct an econometric model of the investment practices of selected Canadian financial intermediaries in order to study the effects that various monetary policy actions have upon their investment behavior. The model incorporates commercial banks, life insurance companies, trust and loan companies, and a private sector. Several types of financial assets are introduced. The results are mixed: most individual equation estimates appear reasonable but the dynamic properties of the model are less satisfactory.

Melvin L. Kliman, Ph.D. Minnesota 1970. Aggregative and sectoral hypotheses on the demand for money: comparative tests for the business sector.

It is argued that money demand hypotheses accepted as useful at the macro level should, under reasonable aggregation assumptions, stand up to empirical testing at the sectoral level. Such hypotheses are tested on data from the *U.S.* manufacturing sector. It is found that neither the permanent income nor wealth functions perform well at the sectoral level, while functions based on theorizing directly related to the business sector provide better results.

Robert D. Laurent, Ph.D. Chicago 1970. Currency transfer by denomination.

This study develops a model relating currency transfers to currency issued and destroyed. The resulting estimates of currency transfers make a significant contribution to the explanatory power of debits in a simple equation of exchange model. The estimates are also used to examine the behavior of currency transfers by denomination, the velocity of currency transfers, and the share of currency in total transfers, over the period 1875–1967. Equations are estimated for currency stock and currency transfers by denomination.

Hsi-Cheng Li, Ph.D. Massachusetts 1971. The economic effects of the proposed federal charter legislation for mutual savings banks.

David E. Lindsey, Ph.D. Chicago 1970. The adaptation of expectations and the consumption function.

A consumption function is proposed and tested in which permanent income is estimated with a variable coefficient of expectations. From Maurice Allais' theory of the demand for money and psychological time, the coefficient of expectations becomes proportional to the desired velocity of money. This theory outperforms the strict permanent income formulation for annual data in the United States.

Raymond E. Lombra, Ph.D. Pennsylvania State 1971. Federal Reserve behavior: identification, rationale, and implications.

Joseph Lucia, Ph.D. Pennsylvania 1970. Central bank policy in Italy.

Josephine M. McElhone, Ph.D. Iowa State 1970. Free gold as a constraint on monetary policy during the early stages of the Great Depression.

Robert D. McTeer, Jr., Ph.D. Georgia 1971. Flexible exchange rates and domestic policy: the Canadian case.

The study examines the ability of flexible exchange rates to insulate the domestic economy from foreign disturbances and to free domestic policy from external constraints. It clarifies and amplifies the theoretical model with respect to economic insulation and independence and finds that Canada's flexible rate from 1950 to 1962 performed well in that regard.

Ahmed A. Malik, Ph.D. Indiana 1970. The money supply process in Saudi Arabia.

The thesis presents a detailed examination of the consolidated balance sheet of the 'Saudi Arabian Monetary Agency,' and a presentation of a consolidated balance sheet for the system as a whole. It carries out aggregation of all these micro relations to form the money supply model for the period considered. A test of the model was performed where the hypothesis was accepted.

Anastasios G. Malliaris, Ph.D. Oklahoma 1971. Monetary growth theory: criticisms and models.

The two main approaches to monetary growth theory are the neoclassical approach and the Keynes-Wicksell approach. The first part of this dissertation argues that both approaches have common methodological errors. The second part of the study is devoted to the presentation of four general equilibrium models, two static and two dynamic, of the introduction of money in a barter economy. All models are solved and the solutions indicate the important role of money in static and dynamic economies.

Lewis Mandell, Ph.D. Texas (Austin) 1970. The demand for money in Israel.

The demand for money in Israel during the period from 1955–67 was found to be a function of permanent income and expected change in prices. Two definitions of money were used in this study. The definition which included passbook demand deposits as well as currency and checking deposits was found to be superior to the definition which did not include passbook demand deposits for prediction purposes.

Douglas L. Maxwell, Ph.D. California (Berkeley) 1971. Measuring tax progressiveness for policy purposes.

Thomas J. Meeks, Ph.D. Duke 1971. A neoclassical theory of monetary equilibrium.

Lorna A. Monti, Ph.D. Texas (Austin) 1970. The demand for money by industry.

A demand for money function for firms is constructed using standard theory of the profit-maximizing firm. Money is regarded as a factor of production.

The three-month Treasury bill rate, a distributed lag on wholesale prices and transactions for the quarter are arguments in this function. The parameters of the function in logarithmic form are estimated by multivariate regression analysis. Large differences among industries are observed.

DONALD J. MULLINEAUX, Ph.D. Boston College 1971. Interest-rate ceilings, dealer transactions charges and minima and noncompetitive bidding for U.S. Treasury bills.

This dissertation examines the effects of regulations concerning the maximum rates that financial intermediaries are permitted to pay depositers on the volume of noncompetitive bidding for new U.S. Treasury bills during periods of high market interest rates. Focus is centered on the extent of small-saver participation in the weekly auction for bills during the intense disintermediation period of 1969 and the initial months of 1970 until the minimum bill denomination was raised to $10,000. The thesis tests a single equation model which explains the level of noncompetitive bidding at the weekly auctions.

CAROLINE A. MURPHY, Ph.D. Clark 1971. Branching and commercial bank performance in New Hampshire: an econometric analysis.

A model is developed to examine bank performance under two legal environments, unit banking and limited branch banking. Selected performance characteristics for each commercial bank in New Hampshire are regressed against structural characteristics of local banking markets for a pre-branching year, 1963, and a branching year, 1968. The performance of branch and unit banks is compared and several explanations are advanced for the less than major impact of branching upon bank performance.

ALAN E. PANKRATZ, Ph.D. Illinois (Urbana) 1971. Quantitative loan ceilings in the United Kingdom: a theoretical and empirical analysis.

This study examines the policy role of quantitative ceilings on clearing bank lending in the United Kingdom. The paper summarizes the rationale for use of this tool, fits it into the larger scheme of policy attitudes and tools, and presents some empirical tests of the impact and effectiveness of loan ceilings in postwar Britain.

MORRIS H. PARDUE, Ph.D. Tulane 1971. An econometric investigation of the supply and demand for bank loans in the United States, 1954–1965.

ELISHA A. PAZNER, Ph.D. Harvard 1971. Optimal resource allocation and distribution: the role of the public sector.

The thesis consists of three independent essays. The first presents an attempt to incorporate the concept of merit wants within the standard model of modern welfare economics. The second derives optimal pricing rules for public enterprises in a world in which the distribution of income is second best optimal. In the third the Averch-Johnson effect is shown to be sensitive to most of the taxes imposed on regulated public utilities, and certain welfare implications are drawn.

A. TIMOTHY PETERSON, Ph.D. Cornell 1970. Monetary instability in the Philippines, 1956–1966.

MANFERD O. PETERSON, Ph.D. Michigan State 1971. A study of asymmetric commercial bank lending behavior.

A multidimensional utility maximization model is adapted to commercial banking. Demands for loan equations for business loans, mortgage loans, and other loans are estimated from weekly balance sheet data for ten large commercial banks. The data covers the period of 1965–1967. The null hypothesis that commercial banks do not adjust their loan portfolios asymmetrically during tight money versus nontight money situations cannot be rejected.

FRANCIS J. PETTIT, D.B.A. George Washington 1970. Effects on selected steel imports of a substitution of a value added tax for the corporation income tax.

This study examines impact of various taxes on international trade in order to estimate effects of a substitution of value added tax for corporation income tax on selected steel imports. Given existing studies of tax shifting and incidence and an examination of 527 transactions involving price quotations for nine imported and domestically produced steel products, the study concludes such a substitution would have little effect on those imports.

EDWARD S. PHILLIPS, Ph.D. Colorado 1970. The allocation of a federal unconditional revenue grant by various techniques and alternative state income measures.

MICHAEL PRELL, Ph.D. California (Berkeley) 1971. Relative movements of *U.S.* price indexes in the postwar period.

JOHN B. REISS, Ph.D. Washington (St. Louis) 1971. A reexamination of the Peacock-Wiseman displacement effect: Pennsylvania as a case study.

Given a level of income and relative prices of public and private goods, a budget constraint is established. A point on the budget constraint is observed. A series of such points are observed over time and a choice path exists when these points are linked. Displacement is a long-term shift in this choice path, shock a short-term shift. Pennsylvania is used to examine how this hypothesis assists identifying the determinants of state spending.

MARY J. RIEDY, Ph.D. Michigan 1971. The impact of restrictive monetary policy upon savings and loan associations, 1965–1967.

HARRISON K. ROBERTS, Ph.D. Indiana 1971. Money demand: a micro-economic study.

STEVEN M. ROBERTS, Ph.D. Purdue 1971. A stochastic inventory model for cash balances with nonzero fixed and proportional opportunity and penalty costs.

This dissertation studies a stochastic inventory model for cash balances using a dynamic programming formulation. Transfers between cash and risk-free earning assets are permissible at the start of each decision period. Sufficient conditions are established which provide a simple optimal policy. A computational algorithm is developed maintaining the dynamic programming features of the model. Examples of multiperiod optimal policies are provided.

ANTHONY M. SANTOMERO, Ph.D. Brown 1971. Optimal transactions behavior and the demand for money.

The thesis develops a fully endogenous model of money demand under certainty. The following results are achieved: i) Through analysis of household behavior, the study determines optimal expenditure flow pattern for each household. ii) Though analysis of firm behavior, the study determines endogenously the optimal frequency of payment. iii) Through market equilibrium analysis, the study determines the effects of changes in transaction costs and rates of return on equilibrium employment and wage rate.

STUART A. SCHWEITZER, Ph.D. Minnesota 1970. Costs and production in banking: the case of the Ninth Federal Reserve District.

This study is an investigation of cost and production functions in banking. A homothetic production function is used to specify the cost function, so that cost function parameter estimates are the basis for production function parameter estimates as well.

BROCK K. SHORT, Ph.D. Cornell 1970. Monetary aspects of the currency board and commercial banks in Malaya and Singapore, 1951–1966.

NANCY D. SIDHU, Ph.D. Illinois (Urbana) 1971. The effects of changes in sales tax rates on retail prices.

The study attempts to locate state sales tax burdens using metropolitan commodity group Consumer Price Indexes. A model is developed in which retail price is a non-linear function of wholesale cost and the applicable sales tax rate. Estimates of a tax-shifting parameter are obtained utilizing fifteen years of quarterly price and tax observations. The results show forward tax shifting in three-quarters of the cases studied and tax absorption in the rest.

THOMAS D. SIMPSON, Ph.D. Chicago 1970. Properties of the demand for consumer durables.

A model of the demand for consumer durables is developed and estimated. Durables are generally found to be more responsive than nondurables to changes in the scale, relative price, and interest rate variables and this is consistent with the theory. In several respects durables are found to be similar to investment. The link between monetary policy and the demand for consumer durables is explored.

JOSEPH F. SINKEY, JR., Ph.D. Boston College 1971. The term structure of interest rates: theory, models of interest-rate forecasting, and empirical evidence.

This study considers two separately conceived models of adaptive forecasting: 1) the Kane expected-change model of interest-rate forecasting (and its restricted reduced-form representation in an expanded error-learning model) and 2) the standard adaptive-expectations model that forecasts interest-rate levels (and its reduced-form embodiment in the Meiselman error-learning model). Within the context of a modified expectations theory, our regression results strongly support the expected-change model over the "level" model.

LEWIS A. SOROKA, Ph.D. McGill 1970. The Canadian personal income tax: the response of revenues to income changes.

DANIEL STEVELMAN, Ph.D. Columbia 1971. Effects of the retail sales tax differential between New York City and the surrounding area, with special reference to relative burdens by income classes and ethnic and racial groups in New York City.

This dissertation deals with evasion of the sales tax in New York City by purchasing outside the city at a lower rate or under no tax. An analysis of location within the city and transportation costs and opportunities reveals that it is the lower income groups, in particular the black and Puerto Rican individuals and families, who pay the highest effective rate of sales tax because of their relative inability to take advantage of lower rates or zero rates outside the city.

JAMES W. THOMSON, Ph.D. Illinois (Urbana) 1970. An empirical investigation of the demand for cash balances by large industrial corporations.

The basic objective was to evaluate several transaction demand models for cash-holding behavior, particularly the Baumol and Tobin transaction demand models, in an attempt to assess corporate motives for holding cash balances. The empirical findings suggest that transaction demand models can account for only a minor portion of the total cash balances maintained by large economic units.

THOM B. THURSTON, Ph.D. California (Berkeley) 1970. Commercial bank portfolio behavior and the commercial loan market.

RALPH H. TODD, Ph.D. Washington State 1971. An empirical postwar study of the short-run shifting of the corporation income tax, 1948–1967.

This dissertation attempts to measure the extent which postwar increases in corporate income taxes have been shifted by *U.S.* corporations primarily engaged in manufacturing. Quarterly data allows one to examine shorter time relationships and, therefore, to obtain more precise knowledge about the speed and extent which the corporate income tax is shifted. This study is limited to the postwar period, 1948–1967.

Salvador C. Umana, Ph.D. Southern Illinois 1971. A study of saving and monetization with special reference to the Philippines.

Part I is a case study of building basic economic statistics on household and corporate savings and financial intermediation in the Philippines from 1951 to 1965. In Part II, a monetization hypothesis of saving (savings per capita and saving-income ratios are some positive function of economic development and of financial development in some decreasing fashion) is given theoretical rationale and empirical content in fifty-five countries.

Gene C. Uselton, Ph.D. Texas (Austin) 1970. Lags in the effects of monetary policy: a nonparametric analysis of the post-accord period in the United States.

David Vanderford, Ph.D. California (Berkeley) 1970. An econometric evaluation of alternative models of the structure of interest rates.

Mary F. Van Loo, Ph.D. California (Berkeley) 1971. The effect of direct investment on Canadian investment.

C. Daniel Vencill, Ph.D. Stanford 1971. Economies of scale in financial intermediation: a statistical investigation of California Savings and Loan Association.

William J. Walsh, Ph.D. Notre Dame 1971. Alleviating the regressivity of the sales tax through vanishing credits.

Steven J. Weiss, Ph.D. Colorado 1971. Uniform formula apportionment of corporate income for state taxation.

Clifford Wheeler, Ph.D. Missouri 1970. Alternative investments for idle government funds.

The two principal alternatives for the investment of idle funds of state and local governments are evaluated in terms of the conditions under which indirect returns on investments in commercial bank time deposits might be realized via operation of an integrated bank credit-regional income multiplier. Empirical findings that pledging requirements adversely affect bank liquidity and inhibit bank lending based upon such deposits suggest the superiority of U.S. Government obligations as outlet for such funds.

Raburn M. Williams, Ph.D. Chicago 1970. Nominal interest rates and aggregate demand: an empirical study of the postwar period.

International Economics

Mohammad Ali-Akbar, Ph.D. Southern California 1970. The welfare cost of protection: a parametric simulation.

This study attempts to demonstrate 1) the hypothesis that tariffs or trade restrictions can entail significant welfare losses through misallocation of resources in production and distortion of prices in consumption, and 2) the hypothesis that uneven tariff rates cause welfare losses of significantly greater magnitude than those obtained when a uniform tariff applies. A general equilibrium analysis previously used by Johnson is used to confirm the plausibility of such hypotheses under a wide range of values of the parameters in the assumed transformation and utility functions.

Ryan C. Amacher, Ph.D. Virginia 1971. International trade, economic reform, and state trading in Yugoslavia.

This study traces the development of Yugoslavia's foreign trade sector and examines the effects of state trading on Yugoslavia's terms of trade vis-à-vis the Comecon nations. Part I of the study traces changes in the commodity composition and direction of Yugoslav trade. Part II reviews the literature on the discrimatory effects of state trading and applies it to the Yugoslav experience. The appendixes include detailed tables of Yugoslav trade movements during the 1960's.

James M. Arrowsmith, Ph.D. Michigan 1970. The International Tin Agreements, 1956–66.

This study examines the major effects of the International Tin Agreements as the basis for an evaluation of their contribution to world economic welfare and the development of the major tin producing countries. The world tin economy is described, the genesis and structure of the Agreements are presented, and an historical account of their operations is given. A theoretical model is developed which explicitly considers private speculation and the simultaneous use of export-control and buffer-stock operations.

Nancy L. Baggott, Ph.D. Purdue 1970. The modern theory of international trade and the Leontief paradox.

Each of the assumptions of the Heckscher-Ohlin theory are analyzed in an attempt to determine which assumptions are "critical" to the theory. A critical assumption is defined as one which cannot be relaxed without reversing the Heckscher-Ohlin prediction statement. Since critical assumptions may be used to explain the Leontief paradox, empirical evidence as to their significance is evaluated in an attempt to determine which assumptions are most likely to explain Leontief's results.

V. N. Balasubramanyam, Ph.D. Illinois (Urbana) 1970. International trade in knowledge: a study of the Indo-foreign technical collaboration agreements in the private sector.

The study analyzes the impact of licensing agreements that Indian firms have contracted with foreign firms over the period 1957–65, on Indian industrial productivity and balance of payments. The interrelationship between Indian imports, foreign direct investment and licensing agreements in various industrial groups is also discussed.

JEAN M. BEILLARD, Ph.D. Notre Dame 1971. Economic integration in the Comecon and the EEC: a comparison.

This study presents the EEC and Comecon as two imperfect economic systems. Systems are characterized by three selected economic networks which provide the basis for comparison in the two blocs. Integration is analyzed via two criteria: (a) the bringing together of national networks; (b) the optimizing process which (a) must bring about. On this basis, it is argued that network imperfections are such that condition (a) cannot be met in Comecon. It is partly met in the EEC, but in a way which is at odds with the conventional theory of customs-unions.

DONALD B. BILLINGS, Ph.D. Oregon 1970. Alternative plans for tariff reduction to favor exports of manufactured goods from the Latin American republics: an empirical analysis.

JOSEF C. BRADA, Ph.D. Minnesota 1971. The allocative efficiency of Czechoslovakian foreign trade.

The allocative efficiency of Czechoslovak foreign trade was investigated by means of a linear programming model which seeks to optimize Czech trade with the *OECD* countries. The model attempts to maximize the difference between the value of Czech imports and exports subject to capacity and payments constraints. The solution of the model indicates that considerable welfare losses occur in Czech trade with *OECD*. Furthermore, the extent of these losses was found to be very sensitive to foreign supply and demand elasticities.

RICHARD A. BRECHER, Ph.D. Harvard 1971. Minimum wage rates and the pure theory of international trade.

The thesis extends the Heckscher-Ohlin type analysis of an open economy to the case where the real wage of a factor is constrained by a floor so that the factor is partially unemployed. Part I sets up the minimum wage model. Comparative statics, emphasizing the response of employment and welfare to specific parametric shifts, occupies Part II. Part III considers optimal and second best commercial policy.

RALPH J. CHARKINS, Ph.D. North Carolina 1970. The theory of economic integration in a development setting: East Africa.

CHAU-NAN CHEN, Ph.D. Chicago 1970. Flexible bimetallic exchange rates and optimum currency areas: a theory of bimetallism.

I-KUEI CHOU, Ph.D. Indiana 1970. Chinese (Mainland) Japanese trade dynamics, 1950–58 with projections to 1975.

The object of the dissertation is to analyze the scope, nature, and development of the economic relation between Communist China and Japan after World War II. Using an econometric model, two projections of China and Japan's trade to 1980 are made in order to examine in detail the potentialities of Sino-Japanese trade over the next decade and in order to examine from the theoretical point of view the implications of this trade. The thesis concludes with a comparison between the Soviet trade patterns in the 1930's and the Chinese patterns in the 1950's and 1960's.

ALAN V. DEARDORFF, Ph.D. Cornell 1971. Growth and trade in a two-sector world.

JOSE R. DE LA TORRE, D.B.A. Harvard 1971. Exports of manufactured goods from developing countries: marketing factors and the role of foreign enterprise.

The need for viable export industries in developing countries is one of the critical questions facing government and corporate planners as we enter the second Development Decade. The central theme of this research is the proposition that the marketing characteristics of a product are major determinants of its export potential from developing countries, particularly as these characteristics affect how, by whom, to what extent, and under what conditions successful export efforts are undertaken.

THOMAS E. DICKINSON, Ph.D. Michigan State 1971. The impact of the United State's changing its tariff on Mexican winter tomatoes: 1967–1977.

Through linear programming analysis this thesis estimates the effect on regional production patterns and trade flows of alternative tariff levels established by the United States against Mexican tomato imports.

JOHN R. DOMINGUEZ, Ph.D. Massachusetts Institute of Technology 1971. Devaluation: its spatial and intertemporal price effects in future's markets.

The principal purpose of this study was to examine the intertemporal and spatial price effects that devaluation may transmit to two closely related markets for the same commodity. Attention was given to explaining very short-run spatial price spread movements during sterling crisis periods in which devaluation expectations may or may not have been present. The analysis showed that the spatial spread may be unresponsive to a sterling crisis or it may behave in a negative manner if expectations foresee alternative corrective measures rather than the use of devaluation.

MICHAEL P. DOOLEY, Ph.D. Pennsylvania State 1971. The Euro-dollar market and *U.S.* monetary policy.

JOHN O. EARLY, Ph.D. Ohio State 1971. An economic analysis of agri-business credit sources and uses in the Itapetininga and Sao Jose Do Rio Preto areas, São Paulo, Brazil.

LOUIS ECKHOUDT, Ph.D. Michigan State 1970. Money and wealth in open and interdependent economies.

Quite recently some authors have questioned the consistency of the traditional macro-economic equilibrium condition: saving equals investment. They have concluded that a "full equilibrium" position can be reached only if saving and investment are both equal to zero. Otherwise, the stock of capital and wealth would still be growing, which indicates that "full equilibrium" is not reached. The thesis studies the implications of this approach in the case of open and interdependent economics, considering not only the exchange of goods but also that of capital.

Milton Esbitt, Ph.D. Michigan State 1970. International capital flows and domestic economic fluctuations: the United States during the 1830's.

This dissertation shows that the American economy in the 1830's and early 1840's was dependent on British cotton and capital markets. Disturbances in these markets led to economic fluctuations in the United States which were intensified by the structure of the American banking and credit system. Contrary to the conclusions of previous works, changes in the supply of specie were shown to be asymmetrically related to changes in the supply of bank money.

Jacob A. Frenkel, Ph.D. Chicago 1970. Money, wealth and the balance of payments in a model of accumulation.

A general equilibrium model of an open economy with money and other assets is developed. The first part investigates the effects of macro-economic policies under alternative exchange rate systems and analyzes portfolio equilibrium and the international distribution of wealth. The second part integrates the monetary and the real sectors in both production and consumption, and develops a theory of the secular evolution of the balance of payments and its components in the process of growth.

Maurice A. Girgis, Ph.D. Indiana 1970. The impact of technological progress on *U.S.* trade.

The main purpose of the study is to theoretically analyze and empirically verify the role of technical progress on the international competitiveness of *U.S.* manufacturing sector during the last decade. To this end, two hypotheses are advanced: a) Industries with high levels of technical progress will show a high level of exports and vice versa; and b) An export industry will have a large share of the international market if it possesses a higher level of technical advance than its foreign counterpart. Various measures of technical progress and trade competitiveness are developed and then examined statistically.

Constantine Glezakos, Ph.D. Southern California 1970. Export instability and economic development: a statistical verification.

Empirical evidence and statistical tests presented in this study support the recently challenged a priori arguments that export instability is generally larger in the less developed countries than in the more developed countries and that this instability is detrimental to the economic growth of the less developed countries only.

Thomas S. Goho, Ph.D. North Carolina 1971. Integration of Latin American chemical industries.

Richard D. Haas, Ph.D. Duke 1971. A portfolio model of international capital flows.

Harriet M. Hinck, Ph.D. Rutgers 1971. Trade diversion in the Commonwealth preference system.

This study develops and tests a model of discriminatory tariff treatment. Since Commonwealth preferences were effected by placing tariffs on nonpreferred suppliers, the static partial equilibrium trade effects are negative trade creation and trade diversion. The variables selected to estimate the trade effects were found to be statistically significant. The study showed that preferences provided an incentive to import commodities at increasing levels of fabrication.

Alfred K. Ho, Ph.D. California (Los Angeles) 1971. Trade liberalization of Japan in the 1960's.

In 1960, Japan embarked on a ten-year plan to double national income and initiated a trade liberalization policy. This study establishes that under certain conditions and in conjunction with certain fiscal and monetary policies, Japan's trade liberalization was consistent with and promoted economic development. Trade liberalization improved imports, facilitated the inflows of capital, capital goods, and technology, and promoted exports. It built up manufacturing and improved the utilization of labor.

Zubair Iqbal, Ph.D. Michigan State 1971. The comparative advantage of developing countries in the manufacturing industries and the effects of generalized tariff preferences.

Under the general preference system, the higher the effective protection in the developed countries, the higher the preference for the developing countries. Therefore, the generalized preferences would give the maximum incentive to the establishment in the developing countries of those industries that are subject to higher effective protection in the developed countries. This study tries to find out as to whether these are the industries in which developing countries have greater comparative advantage and the developed countries, comparative disadvantage.

Rita M. Keintz, Ph.D. Boston College 1971. The demand for international travel to and from the United States.

This study investigates the travel flows to and from the United States on a per country basis within the framework of the balance of payments. Various demand functions are estimated using the least squares method of regression analysis in an attempt to describe the growth of demand and the variations therein. Linear estimates are made for travel to the United States from eighteen countries (European, South American, and the Pacific-area) in both the pleasure and the business categories. Non-linear forms are used for United States travel abroad to six European countries, with demand measured either in terms of the number of travelers or the total expenditures.

Sadikiel N. Kimaro, Ph.D. State University of New York (Binghamton) 1971. International trade theory: treatment of distribution services.

Distribution services—services involved in the actual processes of moving commodities between countries—are studied by: 1) developing a simple model for handling and analyzing intercountry distribution services; 2) interpreting trade relationship between industrialized and raw material producing countries within an analytical framework which includes distribution services; 3) examining possible approaches to empirical studies on international distribution services; and 4) exploring the problem of choosing the country (countries) which efficiently perform the actual tasks of distributing commodities internationally.

PETER A. KOOP, Ph.D. Southern California 1970. Exchange rate policy and export performance.

The hypothesis is that devaluation is a viable development strategy, operating through export stimulation which finances needed capital goods imports. A model is developed which links devaluation and relative export performance on a commodity basis. Export data for a selected group of silver and gold standard countries during the late nineteenth century, when the falling gold price of silver meant continuous, exogenous devaluation in silver countries for more than two decades, provides an empirical test.

WILLIAM E. KOST, Ph.D. Michigan State 1971. Trade flows in the grain-livestock economy of the EEC.

This dissertation develops a production-trade model to simulate changes in international trade for livestock-grain among major EEC countries as a result of formation of the customs union. Estimates of trade with third countries are also made. Short-term projections to 1970 are included.

ARNOLD F. KRONER, Ph.D. Cornell 1971. An analysis of the static welfare effects of a customs union: the European Economic Community, 1958–1966.

MARK L. LADENSON, Ph.D. Northwestern 1970. Balance-of-payments impacts of direct foreign investment of American firms.

In this study we construct a dynamic model in which changes in all sources and uses of funds of foreign subsidiaries of American firms are endogenous. The two such variables which directly affect the capital account of the balance of payments are capital outflows and return dividend flows. We use the model to examine the dynamic response of these variables to a change in the difference between *U.S.* and foreign corporate tax and interest rates.

DONALD E. LEWIS, Ph.D. Washington State 1971. An empirical test of the Heckscher-Ohlin model based on regional trade flows in the United States.

This thesis develops a comprehensive regional test of the Heckscher-Ohlin model using data from the 1963 Census of Transportation. Import-export matrices are prepared which indicate the origin and destination of commodities for the nineteen, two-digit SIC industries. The available evidence on interregional trade flows is inconsistent with the factor proportion hypothesis. Several alternative tests were performed and the hypothesis was consistently inadequate in predicting regional trade flows of manufactured commodities.

ERNESTO S. LIBOREIRO, Ph.D. Michigan State 1970. Effects of the European Economic Community agricultural policies on Argentine exports of beef.

AIDA R. LIBRERO, Ph.D. Minnesota 1971. An analysis of the international demand for Philippine coconut products.

The Philippine coconut market was described by a nine-equation simultaneous model encompassing multimarket and joint product aspects of copra, oil, and meal and the interdependence with other economic sectors. The demands for these products were price inelastic and were significantly influenced by world income and prices of palm kernel, soybean, cottonseed, and sunflower seed oils and soybean meal. The nature and magnitude of market adjustments due to changes in market structure and conditions were also analyzed.

CLAUDIO M. LOSER, Ph.D. Chicago 1971. The intensity of trade restrictions in Argentina: 1939–68.

The study captures nominal and effective protection, using an all-inclusive concept of restrictions to trade. The measurement is pursued considering the difference between internal and external relative prices for the industrial sectors and agriculture. An index of total protection to industry and the social costs of the policies are obtained. It is concluded that the real exchange rate affected incentives. Imports and the growth pattern of the Argentine economy were substantially changed as policies were pursued.

MARTHA F. LOUTFI, Ph.D. California (Berkeley) 1970. An analysis of the cost of foreign aid to the donor country: the case of Japan.

MOHAMED A. L. LOUTFI, Ph.D. California (Berkeley) 1970. Latin American exports of manufactures and semi-manufactures to the United States: an analysis of the effect of unilateral trade barrier concessions.

BARBARA R. LOWREY, Ph.D. Michigan State 1970. The effect of the disparate treatment of tax rebates upon trade patterns of manufactured goods under the General Agreement of Tariffs and Trade.

This study examines whether GATT is theoretically justified in its border adjustment policies and whether the United States suffers under these policies. Border adjustments for consumption taxes are justifiable, but distort trade if imposed at rates equivalent to domestic taxation when external supply or demand elasticities are not infinite. Prohibitions of adjustments for corporate income and employers' social security taxes are unjustifiable. Estimated results suggest, however, *U.S.* trade does not suffer under current policies.

WILLIAM R. LOWRY, Ph.D. Massachusetts 1971. The effects of economic integration in Latin America: 1955–67.

Alan J. MacFadyen, Ph.D. Pennsylvania State 1970. Instability of Canadian foreign trade, 1946–1965.

Donald MacLaren, Ph.D. Cornell 1970. Agriculture and the trade balance of the United Kingdom: a theoretical and econometric analysis.

A model designed to appraise the balance-of-payments effects of alternative agricultural support programs was developed. This model was then used to obtain empirical estimates of the balance-of-payments effects of agricultural policies which have been adopted in the United Kingdom in an attempt to encourage domestic agricultural production and to reduce imports.

John H. Makin, Ph.D. Chicago 1970. The composition of international reserves: a problem of choice involving risk.

Portfolio theory is employed as a means to analyze the choice by central banks between dollars and gold. For nine of the thirteen major central banks considered, the choice between dollars and gold is affected by the rate of return on dollars and the expected rate of return on gold. These results which confirm independent portfolio behavior are consistent with that hypothesis for only a part of the dollar holdings of the central banks. Therefore some degree of interdependence among central banks cannot be ruled out as a factor influencing their choice between dollars and gold.

Lorian L. Marlantes, Ph.D. Stanford 1971. Domestic content requirements and foreign direct investment in developing countries.

Jorge R. Matetic, Ph.D. Minnesota 1971. An economic analysis of the Chilean fresh fruit and vegetable export sector.

Donald J. Mathieson, Ph.D. Stanford 1971. Portfolio balance and international finance.

The first section discusses the influence of portfolio balance in an open outside money economy. The objective is to examine the implications of price flexibility, the government budget constraint, and nontraded goods on the impact of monetary and fiscal policy. The second section examines the implications of a banking system in an open economy. The third section extends the discussion to a growing economy. Finally, the analysis is applied to Germany's balance of payments during the 1960's.

John F. Morrall III, Ph.D. North Carolina 1971. The human capital and product cycle explanation of trade and the *U.S.* demand reversal for human capital-intensive products.

The recent factor proportions and product cycle explanations of *U.S.* comparative advantage were examined theoretically and empirically. It was concluded that the two theories appeared to be equally valid. An examination of the factor proportions assumption that the United States' relatively abundant factor of production is skilled labor revealed, however, that it was true only on a quantity definition and not on a price definition. Evidently, a demand reversal has occurred for skill-intensive products caused by *U.S.* post-World War II national defense and space policies.

John E. Morton, Ph.D. Michigan 1970. Portfolio balance, growth, and long-run balance-of-payments adjustment.

This study is an extension of Mundells' analysis of macro-economic policy and the balance of payments. The Mundell analysis is applicable only in the short run, since the Keynesian model on which it is based defines asset equilibrium in terms of flows rather than stocks and is thus incompatible with conditions of portfolio balance. Existing attempts to correct this deficiency by imposing conditions of zero investment and constant asset stocks are equally untenable as descriptions of long-run equilibrium.

J. G. Guy Motha, Ph.D. Stanford 1971. The impact of import policies on the economy of Ceylon.

This study is concerned with the impacts of changes in import policies in Ceylon in the period 1965–69. An input-output factor constraint model with the base year 1965 is used, and foreign exchange is identified as the binding constraint to further growth of the Ceylon economy. Production responses to changes in administratively determined import allocations in the face of foreign exchange constraint are analyzed to determine effects on particular sectors. Results show that import substituting manufacturing activities were discriminated against.

Yusuke Onitsuka, Ph.D. Chicago 1970. International capital movements and economic growth.

The pattern of economic growth of an open economy is analyzed by considering a small country which can trade capital with perfect international capital market. The various phases of economic growth are discussed according to the capital importer-exporter relation and debtor-creditor relation. The same analytical framework is also applied to the study of the mature creditor (import-surplus) country and to the optimum capital export policy.

Eghosa Osagie, Ph.D. California (Los Angeles) 1971. International reserves and foreign trade adjustment: a case study of Nigeria.

This dissertation compares the cost of alternative policies for dealing with external imbalance in the Nigerian economy between 1950 and 1966. The time-adjusted cost estimates associated with the use of income-changing, flexible exchange rate, and financing policies indicate financing by international reserves as optimal. The rather small and statistically insignificant foreign trade elasticities point to the ineffectiveness of flexible rates. The study suggests international reserve portfolio strategy designed to optimize benefits for Nigeria.

Bodeupdi P. Rao, Ph.D. Ohio State 1970. The economics of agricultural credit-use in southern Brazil.

Don E. Roper, Ph.D. Chicago 1970. A two-country analysis of the distribution of adjustment, capital flows, and monetary policy.

The thesis focuses upon the world money supply and its distribution as influenced by the balance of payments. Several proposals for the international coordination of monetary policies are shown to be essentially the requirement that monetary authorities refrain from sterilization policies. The argument for sterilization policy is related to the determinants of financial capital movements using simulation studies. The distribution of the adjustment burden is related to the choice of intervention currency.

EDWARD L. SAIERS, Ph.D. Pennsylvania State 1970. The instability of export proceeds of selected sub-Sahara African countries.

SALVATORE SCHIAVO-CAMPO, Ph.D. Columbia 1970. Essays in export instability.

The first essay surveys instability indices and describes the more efficient ones. The second analyzes international instability patterns: while overall instability declined in recent years, the difference between *LDC*'s and *DC*'s increased markedly. The third surveys causes of instability. The fourth finds an inverse association between country size and instability. The last finds a direct association between export instability and instability of developmental imports, and between the latter and instability of *GDP*.

TODD SANDLER, Ph.D. State University of New York (Binghamton) 1971. A contribution to the theory of devaluation.

The dissertation builds upon a synthesis similar in structure to the model of Tsiang. Capital flows are added onto the model transforming it into a balance-of-payment model. In addition, the paper contains a comparison between the synthesis model in real-terms versus the model in money-terms. A two-country Samuelson investment accelerator model is derived in the later part of the dissertation. Also, a speculative spending accelerator is formulated which expresses the dynamic price expectation effect of Alexander.

WILLY SELLEKAERTS, Ph.D. Michigan State 1971. The effects of the EEC on East-West trade.

This thesis is a study of the trade expanding and diverting effects of the EEC on her members' imports from the Communist countries of eastern Europe. The empirical results indicate that Romania, Bulgaria, Yugoslavia, Albania, Hungary, the GDR and Poland shared in the extra-area trade expanding effect of the EEC. The USSR and Czechoslovakia, on the contrary, suffered from trade diversion. As a group, the Communist countries of eastern Europe were favorably affected by the formation of the EEC.

JOSEPH SHAW, Ph.D. California (Los Angeles) 1971. The British convertibility "plan": a study in policy evolution.

This dissertation is a history and analysis of British foreign economic policy between late 1951 and mid-1955. It studies efforts by the government to achieve its major foreign economic policy goal, viz., the restoration of nonresident current account convertibility of the pound sterling and the simultaneous initiation of a floating exchange rate. It discusses the events of the period and treats at length the implications of convertibility and a floating rate for the British economy and for western economic institutions.

EDWARD G. STARR, Ph.D. Minnesota 1971. The cost of import-substitution: the case of the Indian caustic soda industry.

The goals of import substitution are to reduce dependence on foreign suppliers and to develop domestic industries. By these criteria the Indian caustic soda industry is "successful." If maximization of the net international value added of goods and services produced by a given set of domestic resources is the goal, import-substitution in caustic soda is an inefficient use of India's resources. For 1966, the estimated costs of this policy was $15–20 million.

WILLIAM E. STEELE, Ph.D. New York 1970. The role of nonprice competition in the analysis of international trade patterns.

Using a simple statistical model, this paper tests empirically Staffan B. Linder's hypothesis that trade flows between countries will be greater the more similar their demand structures. This hypothesis is contrary to the traditional theory which implies that trade will be greater between countries the more dissimilar their demand structures are. The test results were inconclusive.

AKRAM M. STEITIEH, Ph.D. Ohio State 1971. Input productivity and productivity change of the crop enterprise in southern Brazil.

DOUGLAS STEWART, Ph.D. Oregon 1970. Intermediate goods in trade: implications for pure trade theory.

ARNOLD H. STUDENMUND, Ph.D. Cornell 1970. International trade, economic development and the incremental capital-output ratio.

FELIPE SUVA MARTIN, Ph.D. Massachusetts Institute of Technology 1971. *U.S.* direct investments in the Philippines.

IGUSTI B. TEKEN, Ph.D. Purdue 1971. Supply of and demand for Indonesian rubber.

The objective of this study is to determine the nature of supply and demand relationships in the Indonesian rubber industry and to analyze the policy implications of selected key variables. A simultaneous equations model consisting of six behavioral equations and three definitional identities was developed to describe the annual structure of the Indonesian rubber industry. Three-stage least squares procedures were used to estimate the parameters of the behavioral equations in the model.

G. RICHARD THOMAN, Ph.D. Fletcher School 1971. A theory of investment incentives: foreign investment and Belgian regional development.

Edward Tower, Ph.D. Harvard 1971. Financial policy under alternative exchange rate systems.

The first part of this thesis is a critical survey of earlier contributions to the study of monetary and fiscal policy under fixed and flexible exchange rates. The second is a reformulation in the light of recent theoretical developments of certain aspects of the models conventionally used on these studies. The third part uses a special case of the reformulated model to analyze the effects of monetary and fiscal policy under the two exchange rate systems.

Henri Vandendriessche, Ph.D. California (Berkeley) 1970. The effects of budgetary policy on economic growth and instability.

Jacinto Vereda, Ph.D. Pennsylvania 1971. Structure and export competitiveness of the Spanish wool industry.

The Spanish wool manufacturers can compete internationally. Their difficulties lie more on the marketing side than on the production side; their character of residual suppliers still hinders their effectiveness. Due to the nature of the products and to the production process itself, small firms are needed, but they should cooperate in order to overcome the many shortcomings presented by small size.

Percy D. Warner III, Ph.D. Michigan State 1970. The impact of the service on the foreign debt on the monetary structure in Argentina: 1955–1965.

This dissertation tests the hypothesis that foreign debt service has been a major deterrent to economic development in Argentina particularly, but not exclusively, through effects on the internal monetary structure. The results of the analysis suggest that the hypothesis is invalid.

James A. Wilson, Ph.D. Wisconsin (Madison) 1971. Trade and factor growth.

The thesis presents a dynamic version of the Hecksher-Ohlin model in which 1) all factors of production are a form of capital and 2) investment in each factor and changes in relative factor endowments are related to relative factor prices. Without knowledge of factor supply elasticities, the international division of production is ambiguous when both trade and factor growth occur simultaneously. Given a relatively high supply elasticity for physical capital, a possible explanation of Leontief's paradox is offered.

Warin Wonghanchao, Ph.D. Cornell 1970. Impact of foreign trade and government policies on economic growth and development in Thailand, 1955–1967.

James G. Wrone, Ph.D. Illinois (Urbana) 1970. Effective rates of protection in the United States: an analysis of the Kennedy round tariff reductions.

Chung Chi Yu, Ph.D. Catholic 1971. Theory, controversy, and empirical study of arbitrage, speculation, liquidity and short-term capital movement, from 1958 to 1969.

This research is intended to test the arbitrage theory and to restructure the theory of speculation and liquidity in relation to short-term capital movement, together with testing the validity of the revised theory. The operations of the Euro-dollar market and the gold arbitrage are also analyzed. This study reveals that factors affecting the short-term capital movement are 1) the spot speculative expectation, 2) the excess of the domestic liquidity, and 3) the external deficit of the United States.

Business Administration; including Business Finance and Investment, Insurance, Marketing, and Accounting

Moustafa H. Abdelsamad, D.B.A. George Washington 1970. Evaluating capital expenditure proposals in large industrial corporations.

This dissertation reports the methods used by large industrial corporations in the United States in evaluating capital expenditure proposals. The data were obtained by means of a questionnaire sent to firms selected from the largest 500 industrial corporations listed in the June 1968, *Fortune Directory*. The topics covered include: selected administrative aspects, current methods, appraisal of methods by the respondents, common problems, and selected techniques for capital expenditure evaluation.

Emil Albert, D.B.A. Michigan State 1971. Aluminum ingot market: 1959–1968.

The bases of the study are trend analyses of ingot deliveries, shipments, production, and inventories. Causes of fluctuations in the market are identified from historical evidence and an industry survey. Effects of the fluctuations are compared with theories relating to constant stock-sales ratios, price speculation, and hedging against supply interruptions. Results indicate that market fluctuations are the result of other than changes in demand and that the constant stock-sales ratio is not valid.

Lane K. Anderson, Ph.D. Wisconsin (Madison) 1970. The framework of an information monitoring system for an Economic Order Quantity (*EOQ*) inventory model.

The hypothesis underlying this study is that the framework for an information system which is centered on an *EOQ* inventory model and based on current or expected cost data can be developed for monitoring 1) the validity of estimated (standard) values used in the components of the *EOQ* model and its related order point, and 2) whether actions specified by the *EOQ* model and its order point are performed within the limits established by management.

Said M. Arafa, Ph.D. California (Berkeley) 1970. Accounting information for planning and control models: the case of pert and critical path method.

ROBERT H. ARNOLD, Ph.D. Northwestern 1970. An examination of the needs of terminal decision units: a management information systems approach.

This study grew out of the recent trend toward closely coupled, man-machine decision systems, made possible by the advent of online computer systems. The results of the research presented in the dissertation suggest that there will be important variances in data requirements between terminal and nonterminal decision units in similar decision situations. These findings are discussed in terms of the use of formal models, search, problem formulation, and the integration of organizational systems.

SABER A. M. AWAD, Ph.D. Michigan State 1970. The acceptance of contemporary management accounting techniques in small manufacturing firms.

The objectives of this research have been to find out the extent of application of modern management accounting techniques in current practice of small and medium manufacturing companies; to investigate the diffusion of management accounting innovations; and to identify the environmental conditions which influence this diffusion process. Modern management accounting techniques are not widely practiced among small and medium companies. Top management's attitudes and the management accountant's attitudes toward change and his professional competence are the most important factors influencing accounting innovations.

PETER M. BANTING, Ph.D. Michigan State 1971. Imagery in vertical distribution channel structure: an empirical investigation of steel service center images.

Images of *U.S.* steel service centers perceived by steel mills, service centers, and steel buyers are examined. Using semantic differential, factor analysis, and multiple regression techniques, similarities between images projected and perceived are identified; key elements representing the image spectrum, ranging from unfavorable to favorable, are isolated; a basic operational tool for image measurement by centers is developed; areas offering opportunities for image improvement are identified; and the image-profit relationship is tested.

ELBA F. BASKIN, JR., Ph.D. Michigan State 1970. An empirical evaluation of the communicative effectiveness of consistency exceptions in auditors' opinions.

The purpose of this research effort is to examine the information content of consistency exceptions in auditors' opinions and to provide empirical evidence which will support a judgment as to the effectiveness of the consistency exception communication network. Also, certain questions concerning the concept of materiality are explored. Based on statistical results, the conclusion is reached that consistency exceptions appear to possess no information content. Also, investors do not appear to possess materiality functions.

WILLIAM M. BASSIN, D.B.A. George Washington 1970. An empirical evaluation of methods for estimating the parameters of Weibull restoration process hazard functions, and for determining optimal overhaul intervals.

If equipment failures follow a Weibull restoration process hazard function, there are several methods for estimating the hazard function parameters and several optimal overhaul interval models. Both the estimation methods and optimality models are evaluated by simulation. The results indicate that the maximum likelihood and minimum risk estimators are the most useful. The most effective model chooses an optimal time between overhauls. Bayesian models suffer from parameter estimation bias.

WALDRON BERRY, Ph.D. Florida 1970. A planning program budget system for a university student health service.

JOHN G. BEVERLY, Ph.D. North Carolina 1971. Using quantitative methods to develop a flexible credit limit policy for a small manufacturing firm offering trade credit.

Into eight quarterly (seasonal) time-cross section stepwise multiple regressions, forty independent variables derived from the customer ledger sheet were entered, with the result that ten or less explain more than 50 percent of expected net present value ratio (time-adjusted rate of return) on credit sales. Posterior distributions of ENPV-Ratio-Predictors being constructed via computer card image sort, a reverse-induction sequential-sampling credit management model predicts individual customer profitability, incorporating cost of information.

JAMES D. BLUM, Ph.D. Michigan State 1971. An analysis of the price behavior of initial common stock offerings.

The market price behavior of initial common stock offerings was studied to determine how they have performed in the past. The major findings of the study were: Initial common stock offerings have shown higher than normal market rates of return; initial offerings' market price performance has been in disequilibrium with the over-the-counter average; initial primary common stock offerings have proved to be a better investment than the other types of initial offering distributions (secondary and mixture offerings).

LAWRENCE P. BOHL, Ph.D. Purdue 1971. An analysis of the behavior of firms supplying purchased inputs to agriculture: an experimental approach.

This study examined the influence of organizational structure, agricultural environment, and competitive prices on the behavior of farm input supply firm managers. A business management game, specifically developed to simulate the farm supply industry, was used to measure the behavior and performance of managers under controlled behavioral laboratory conditions. A nested factorial ANOV model was utilized to test the influence of three main effects. Multiple regression was used to test relationships between experimental data and other covariates.

G. Geoffrey Booth, Ph.D. Michigan 1971. The management of a commercial bank portfolio: a programming approach.

Robert Briscoe, D.B.A. Harvard 1971. Traders and idealists: a study of the dilemmas of consumers' cooperatives.

This study attempts to increase our understanding of some of the reasons for success and failure in consumers' cooperatives. In particular, it sets out to demonstrate that the conservatism which characterizes the behavior of so many co-ops stems, in part, from the active participant's ways of perceiving the world of business and from the disabling dilemmas he experiences when trying to reconcile his social ideals with the day-to-day running of a supermarket.

Max Brown, Jr., Ph.D. California (Berkeley) 1971. An investigation of multiple-product firms.

Robert L. Brown, Ph.D. Florida 1970. Competition and the Florida automobile insurance plan.

The Florida Legislature passed an open competition rating law in 1967, replacing a prior approval law. The resultant competition has changed the market for automobile liability insurance. This study demonstrates by statistical analysis the favorable decline in assigned risks, substandard risks, and addition of new standard insurers. The conclusion is drawn that the motorist, the insurers, and insurance regulating agency have benefited from this viable approach to the rating problem.

Richard W. Brunson, Sr., D.B.A. Michigan State 1970. A behavioral case study of a mini-conglomerate.

This research is believed the first intensive comprehensive case study of a mini-conglomerate. A detailed questionnaire was responded to by 113 members ranging from foremen to president. Organizational effectiveness (criterion) was found to be significantly related to four other areas (predictors) measured: personal values of "membership," "organizational" (structural), "dynamic" (process), and external environment. The contingency theory of Larsch, Lawrence, and Morse appeared to be supported while the Likert participative theory appeared not to be fully supported.

Jimmy C. Caldwell, Ph.D. Alabama 1970. An inquiry into business gaming as a pedagogical technique in accounting education.

Elias R. Callahan, Jr., Ph.D. Alabama 1970. The development of an optimization technique for general stochastic inventory simulation models.

Charles R. Carlson, Ph.D. Michigan State 1971. A financial efficiency model.

A model, using nonparametric statistics, was developed for a firm in a given industry to assist in establishing long-range financial guidelines. Specifically, an index of financial efficiency was used in determining the appropriate level of current ratio, cash, inventory, fixed assets, long-term debt and cash dividends. The data for eighteen drug companies for 1960–69 were obtained from Compustat. Weighting factors indicating the relative importance of several financial ratios to shareholder wealth maximization were included.

R. Wayne Cartwright, Ph.D. Purdue 1971. Research management in a department of agricultural economics.

A preliminary appraisal of the applicability of selected quantitative techniques to research management in academic agricultural departments. The focus of the study was a single case department of agricultural economics. Mathematical models were proposed for the selection of research areas and research job portfolio selection. Software for a research monitoring and reporting system was constructed and was demonstrated to be semioperational in the current environment.

Derek F. Channon, D.B.A. Harvard 1971. The strategy and structure of British enterprise.

This thesis describes the growth and development of the largest 100 British manufacturing corporations over the period 1950–1970. The strategic response of most has been diversification in response to changes in competitive and market forces. Changing strategy has led to the widespread adoption of a multidivisional organization structure which differs especially from that found in the United States in the lack of incentive reward systems. A number of integrated industries were identified which had not diversified.

Mo-Hung Che, Ph.D. Cornell 1971. A treatment of distributed lags in the measurement of advertising effectiveness.

William O. Cleverley, Ph.D. California (Berkeley) 1971. Systems cost analysis for hospitals.

Sheldon N. Cole, Ph.D. Michigan 1971. Identical vs. distinct concepts of income for federal income taxes and financial reporting.

Henry H. Conner, D.B.A. Harvard 1971. Financial planning and control education for engineering managers.

The American Society for Engineering Education and the Engineers' Council for Professional Development recently formally recognized the need for management education for engineers. The purpose of this research is to identify the relative usefulness of the elements of financial planning and control education to engineering managers.

Charles J. Constable, D.B.A. Harvard 1971. An evaluation of computer process control in the British steel industry.

The thesis considers the results achieved at thirty-seven on-line computer process control installations in the British steel industry. Data were collected from twenty separate locations and over 100 people closely associated with the installations were interviewed.

James L. Cottrell, Ph.D. Pennsylvania 1970. An environmental model for performance measurement in a chain of supermarkets.

Environmental variables which are significant estimators of a supermarket's behavior are combined into a "least squares" planning and control system that will pinpoint deviations from expected performance in period $t+1$. Performance is evaluated based on forecasts of sales and gross margin percentages by product line and also estimates of controllable expenses. The effectiveness of the unit manager is assumed to be in the residual and control limits are set at $\hat{y} \pm 2\sigma_{\hat{y}}$.

Alice E. Courtney, D.B.A. Harvard 1971. Communications research and advertising practice.

The thesis describes the advertising communications process as viewed by practitioners in client organizations and advertising agencies. Personal interviews were carried out with sixty-six practitioners. Each traced through the development of an advertising program with reference to a specific communications situation. The research identifies the decision variables practitioners consider, and pinpoints those areas where communication research is relevant to the practitioner. The state-of-the-art in communications research is evaluated with reference to practitioners' needs. Priorities for future research are suggested.

Terry L. Crawford, Ph.D. Cornell 1970. Managing a data information system.

A system of collecting and distributing market information throughout a seven state apple producing area was designed and operated. The system used the *U.S.* mail, mark-sense cards, paper tape, teletype, and a computer. Estimation procedures were developed and tested. Information was computer prepared in tabular and graphic form. It was concluded that currently available equipment offered great flexibility and that costs of such a data system is not prohibitive. Estimation procedures gave reasonable accuracy.

Gary M. Crooch, Ph.D. Michigan State 1970. An investigation of investors' financial statement knowledge.

This research is an attempt to measure the knowledge of financial statement terms and concepts currently held by the nonprofessional investor. Further, the research attempts to discover the relationship between financial statement knowledge and certain demographic, behavioral, and attitudinal variables. The research methodology employed the use of a pretested, direct-mail questionnaire. A representative conclusion of the research is that the average nonprofessional investor does not have sufficient knowledge of financial statements to use them in an informed manner.

Ronald C. Curhan, D.B.A. Harvard 1971. A study of the relationship between shelf space and sales for selected products in self-service food supermarkets.

The impact of shelf space changes on unit sales was examined for over 400 grocery products in four large self-service food supermarkets under actual operating conditions. Coefficients of "space elasticity"—the ratio of relative change in unit sales to relative change in shelf space—were calculated for each product. A model was proposed wherein space elasticity was hypothesized to be a function of a number of product-specific variables including physical properties, merchandising characteristics, and use characteristics.

Henry L. Custer, Ph.D. Alabama 1970. The courts' concepts of accounting principles as revealed by court decisions.

William W. Damon, Ph.D. Cornell 1971. A model for simultaneous decision making among functional areas of the firm.

Heinz Dinter, Ph.D. Florida 1970. A theory of computer based management control.

Maynard M. Dolecheck, Ph.D. Alabama 1970. A study and quantification of executive attitudes and preferences pertaining to decisions involving risk.

Roger L. Dunbar, Ph.D. Cornell 1970. Some characteristics of the interaction between budgets and people.

John W. Duncan, D.B.A. Harvard 1971. The beginning management consultant: a study of individual adaptation to job complexity.

This thesis presents the results of research on seventy-one recent graduates of Harvard's M.B.A. Program who entered the field of management consulting. The primary objective of the thesis was to explore how and why these individuals felt they were adapting to their jobs as general management consultants.

Thomas F. Funk, Ph.D. Purdue 1971. Market analysis of branded commercial feeds.

The underlying objective of this study was to determine the characteristics of the market for branded commercial feeds which influence the development of marketing strategies by commercial feed firms. Specific objectives were: 1) to estimate the mixed feed demand functions facing individual firms; 2) to investigate competitive patterns of the mixed feed market; and 3) to investigate farmer purchasing behavior characteristics. Multiple linear regression models and a Markov brand switching model were the major analytical techniques used.

Asa B. Gardiner, Ph.D. Florida 1970. Corporate planning in Florida banks.

The level of corporate planning in Florida banks as reflected in deposit growth June 30, 1961 to June 30, 1966 was measured comparatively. Questionnaire responses from 151 banks were converted into a novel Planning Level Indicator. Forty-four banks, planning above average, showed a very significant 28.7 percent difference in deposit growth for the period compared to thirty-two below-average planners. The differentials had increased to 38.0 percent by June 30, 1969.

Arne Gausel, D.B.A. Harvard 1971. Determinants of consumer brand switching: an experiment.

Consumer brand switching was conceptualized in a risk-taking framework, primarily as a function of perceived risk on the one hand and risk-taking propensity on the other. A small-scale laboratory experiment was conducted, in which paper-and-pencil choices among leading brands of ground coffee and toothpaste (at varying prices) served as a surrogate for actual buying behavior. Perceived brand differentiation was by far the strongest switching determinant, followed by product importance and deal proneness.

Peter Gilmour, Ph.D. Michigan State 1971. Development of a dynamic simulation model for planning physical distribution systems: validation.

This dissertation is a formal study of validation of computer simulation models in general, and in particular an analysis of the performance of a model of an integrated physical distribution system. A general validation procedure is developed testing long-term stability, predictive ability, and the sensitivity of the assumptions of a model. Many statistical techniques are used for this testing with most emphasis placed on spectral analysis.

Charles B. Handy, Ph.D. Iowa State 1970. The accounting for convertible debt.

An investigation into the reasons underlying demand for and supply of convertible debt. The findings were used to determine value of the conversion privilege along with the accounting procedures to be followed on the books of the issuing company.

Harold M. Harris, Jr., Ph.D. Purdue 1971. The economic outlook for the independent dairy company.

The objective of this study was to determine whether independent dairy companies can be expected to survive and remain a viable, competitive force in our economy. Data were gathered mainly from detailed interviews with executives of twenty large representative case firms. Supplemental information from many other sources was also gathered. The outlook for this institution is not outstanding, but there are possibilities and good opportunities for well-managed firms of this type.

Margaret M. Hennig, D.B.A. Harvard 1971. Career development for women executives.

This thesis describes an in-depth clinical field study of the life and career histories of twenty-five top level women executives. Its overriding objective was to discover what it was about these women—their backgrounds, their education, their personality development, their social experiences, their job experiences and their resulting self concepts, self ideals and behavioral styles—that made it possible for them to succeed in administrative roles which their culture and their society generally considered almost totally masculine.

Regina E. Herzlinger, D.B.A. Harvard 1971. Control of not-for-profit organizations: an experiment. The implementation and evaluation of a management control system in the Bunker Hill Health Center.

The purpose of this thesis was to evaluate the impact of installing a management control system, of the kind used in the private sector, on the effectiveness and efficiency of a not-for-profit medical organization. To this end, a management control system was installed in the Bunker Hill Health Center in the period from January to April 1970 and its impact on the Center was observed from May to October 1970.

Yusuke Horiguchi, Ph.D. Rice 1971. Corporate financial decisions: a general equilibrium approach.

The development and application of a general equilibrium model with special emphasis on the interdependence of the firm's investment, financing and dividend decisions casts a new light on various aspects of corporate financial behavior. We derive an investment theory which unifies assorted past empirical findings into a comprehensive theoretical construct. Our realistic solutions to the problems of optimal financing and dividend payout are a new challenge to the Modigliani-Miller position.

Hugh A. Hoyt, Ph.D. Michigan State 1970. The relative predictive capacity of two bank earnings measures: an empirical evaluation.

The purpose of this study was to determine the relative capacity of two bank earnings measures to predict future all-inclusive earnings. The two earnings measures used were Net Operating Earnings, the measure preferred by bankers, and Net Income, the measure preferred by accountants. The two earnings measures differ in that Net Operating Earnings contains no provision for loan losses and no security gains and losses. Neither measure was found to have a superior predictive capacity.

James P. Hynes, Ph.D. Michigan State 1971. Truckload motor carrier rates in a normative spatial environment.

Spatial effects in truckload freight rate structures are examined in this dissertation in terms of a profit maximizing motor carrier. In general, a rate structure will emerge which is polarized with respect to each terminal in the system. These polarized patterns were found to exist in actual general commodity freight rates of a motor common carrier.

Carol Inberg, Ph.D. California (Berkeley) 1970. The general accounting office: a case study in the audit techniques as a means of determining management's effectiveness in carrying out an organization's policies.

Bob M. Inge, D.B.A. Harvard 1971. The influence of fund sources on financial decisions in nonprofit hospitals.

The focus of this study is on how financial decisions are made within nonprofit hospitals and how these decisions affect costs. From a narrow point of view, the study primarily describes and analyzes the influence of fund sources on the costs of operating the clinical laboratory, the diagnostic radiology, and the dietary services in two Veterans Administration, two Army, and two Massachusetts community hospitals.

MORTON I. JAFFE, D.B.A. Harvard 1971. The decision process in advertising media selection.

The primary thesis research problem and objective has been to develop an (improved) "normative but practical" conceptualization or model of the media decision-making process focusing mainly on the concepts and major operational procedures which advertisers and their advertising agencies should use to make decisions on the selection and use of advertising media.

MAHAVIR JAIN, D.B.A. George Washington 1971. An analysis of price change associated with stock splits.

This dissertation examined the relationship between stock splitting and market price change. The price change was measured between two dates: (1) eight months before the split date, and (2) four months following the split date. A multiple regression model was used and such variables were included as earnings, dividends, market trend, earnings growth, and the split factor (dummy variable). It was concluded that there was no statistically significant price change associated exclusively with stock splitting.

THOMAS E. JOHNSON, JR., Ph.D. Alabama 1971. A comparative benefit-cost analysis of industrial development bonds and federal tax credits as tools for stimulating regional economic development.

CHARLTON H. JONES, Ph.D. Michigan 1970. An empirical development of a linear work load model for faculty course instruction.

MADHAV P. KACKER, D.B.A. Michigan State 1970. An exploratory study of the marketing adaptation of *U.S.* business firms operating in India.

The study explored the nature and patterns of marketing adaptation or adjustments of American business firms operating in India. It was found that these business firms adapted their marketing effort to a limited extent in response to various environmental forces. They showed a greater amount of voluntary adaptation in the areas of promotion and distribution than in those of product and pricing. Most of them were interested in maintaining their international product standards, yielding only to mandatory changes in product design and packaging.

MOJTABA KASHEFI-ZIHAGH, Ph.D. Michigan State 1970. An empirical investigation of the relationship between value systems and organizational effectiveness.

The organizational and management school did not devote any attention to the value systems of organization members as a probable factor for organizational effectiveness. The study analyzed the value systems of 245 executives from the most and least effective suborganizations of a bank, an insurance company, and an educational institution. The effective executives had a value system significantly different from that of the ineffective ones. They emphasized competence over moral values and effectiveness over profitability.

FREDERICK D. KENAMOND, Ph.D. Alabama 1971. Accounting theory and practice, social-economic-political environment and the federal income tax: their interrelationships concerning inventories.

JOHANNES KINFU, D.B.A. Michigan State 1970. The accounting provision of the commercial code of Ethiopia.

The study describes the accounting and auditing provisions on share companies (corporations) of the Ethiopian Commercial Code promulgated in 1960. Recognizing the importance of these provisions for providing the accounting information indispensable for adequate financial planning and control of enterprise efficiency, and aiding the overall planning of economic development, the study examines the underlying rationale behind the code's accounting provisions. It evaluates their relevance, appropriateness and adequacy of full disclosure in providing information for sound economic decision making.

THOMAS P. KLAMMER, Ph.D. Wisconsin (Madison) 1970. A study of the association of capital budgeting techniques with firm performance and firm characteristics.

An empirical study of existing capital budgeting techniques involving both published firm financial data and questionnaire responses. Some answers are suggested for the following questions. What capital budgeting techniques are being used by firms today? Is there a pattern of capital investment procedures associated with characteristics of classes of firms such as capital intensity or risk? Is a firm's performance related to the sophistication of the capital investment procedures and standards it uses?

LARRY F. KONRATH, Ph.D. Michigan 1970. Accounting and auditing methods used in the prevention and detection of employee fraud.

MARVIN R. KOPULSKY, Ph.D. Northwestern 1970. Impact of tax practices on financial reporting.

Generally accepted accounting principles today consist of a variety of allowable alternative methods. Thus, the accountant finds himself in an environment that forces him to make decisions regarding the choice of one of several permitted procedures for reporting items on financial statements. An understanding of financial reporting practice requires knowledge of factors that influence such decisions. The purpose of this paper is to critically examine one such influence: income tax considerations.

GEORGE W. KRULL, JR., Ph.D. Michigan State 1971. A study of two-year college transfer students' elementary accounting achievement.

The purpose of the study was to determine whether two-year college transfer students and university nontransfer students enrolled in intermediate level accounting differed significantly with respect to their levels of achievement in elementary level accounting. The major findings were that 1) two-year college transfer students had a lower degree of elementary accounting achievement than nontransfer students, and (2) two-year college students did not have similar achievement to nontransfer students in intermediate level accounting.

HERBERT V. LADLEY, D.B.A. George Washington 1970. A business simulation study of the behavioral implications of price level adjustments to financial statements.

This study combines the controversy regarding adjustments to financial statements for the effects of inflation with the relatively unexplored subject of the behavioral effects of alternative accounting methods.

JOSEPH C. LATONA, D.B.A. Kent State 1970. The influence of certain external variables, functions and behavior on organizational effectiveness.

This exploratory study revealed that differences in performance levels that exist between organizations within the same area of endeavor may be attributed to one or a combination of the following factors: contrasts in 1) external variables, 2) the manner in which certain functions are performed by each organization, and 3) certain internal variables such as behavioral patterns (as measured by individual behavior, intragroup behavior, lateral intragroup behavior, vertical and intergroup behavior) of its members.

GEORGE E. LEE, Ph.D. Purdue 1971. Exploitation of information for capital accumulation under uncertainty.

The interrelationships between predecisions, solution decisions, and value of solutions were examined. This was done by analyzing the problem of predetermining the length of the planning horizon for capital accumulation on a case farm. Integer solutions under uncertain product price ratios for various planning horizons were developed, using a synthetic allegative solution procedure. A short planning horizon sufficed to analyze the labor-limiting case, whereas a longer one was required for the capital-limiting case.

STEPHEN E. LOEB, Ph.D. Wisconsin (Madison) 1970. A behavioral study of CPA ethics.

The study examined the ethical behavior of certified public accountants in their professional roles and identified the factors that affected conformity to or violation of professional norms. The formal control system that exists in a large public accounting community was described and analyzed. A determination was made of the degree of compliance to the official norms of accounting practiced by the members of this community.

PETER M. LYNAGH, Ph.D. Michigan State 1971. Physical distribution patterns for the metal service center industry.

This thesis studies physical distribution patterns in the Metals Service Center Industry. The specific purposes of this research were to: 1) compare the existing patterns with a maximum performance model; 2) determine if size affects the relationship of a Service Center to the model; 3) determine if profit affects the relationship of a Center to the model; and 4) compare differences of opinion regarding customer service among those holding different jobs within the Center.

JOSEPH A. MCHUGH, Ph.D. Michigan State 1970. The derivation of learning task hierarchies and instructional objectives for training United States Air Force base-level responsibility center managers and resource advisors.

This research employs a newly developed model for the derivation of instructional objectives for base-level resource managers and their resource advisors. These products should provide the foundation of an effective resource management training program to be applied throughout the United States Air Force. They also represent inputs to the remainder of the instructional process, i.e., the specification and development of actual learning experience.

EDWARD V. MCINTYRE, Ph.D. North Carolina 1971. An empirical study of the effects of current cost financial statements on investment decisions.

In this study undergraduate and graduate business students were asked to make decisions concerning the desirability of investing in the common stocks of two pairs of companies. The purpose of the experiment was to see if subjects using current cost financial statements made decisions that were different and better than those using historical cost statements. No significant differences could be detected in the decisions of the subjects, regardless of the kind of information they received.

DAVID B. MACKAY, Ph.D. Northwestern 1971. Consumer movement and store location analysis.

The study developed a dynamic, sequential decision process model of individual shopping behavior. Stochastically defined by sets of discriminant functions, the model simulates the multi-stop movement and expenditures of consumers. Primary contributions include 1) the development of a methodology which effectively models multi-stop movement and 2) the incorporation of attitudinal factors in a spatially defined location analysis model. The model, which predicts the sales of future stores, was empirically validated.

PATRICK B. MCKENZIE, Ph.D. Michigan State 1970. The relative usefulness to investors of price-level adjusted financial statements: an empirical study.

The purpose of this dissertation was to empirically investigate the relative usefulness (to investors) of 1) financial statements adjusted for changes in the general price level and 2) conventional adjusted financial statements. Usefulness was defined as predictive ability. A multiple correlation model employing selected financial ratios as inputs and a simple forecasting model used to predict future income values provided the needed empirical evidence. In general, no significant differences were observed in the usefulness of the two sets of financial statements for the periods studied.

THOMAS E. MCLEOD, Ph.D. Alabama 1971. An evaluation of services to management in connection with the annual audit.

FRANKLIN R. MADERA, D.B.A. Harvard 1971. Measurement, analysis and reporting of performance in multidimensional corporations.

Translation of foreign currency financial statements has been traditionally viewed as a historical process. This study suggests that a significant contribution in the management of foreign operations will be achieved by adding a forward looking approach. Recognition of the behavioral implications of the translation process will have positive impact upon control systems for foreign operations. The selection and implementation of a translation method appropriate for the environmental conditions of each foreign operating unit and for the overall objectives of the multinational firm will enable more effective communication between headquarters management and the manager in the field.

EDWARD J. MARIEN, Ph.D. Michigan State 1970. Development of a dynamic simulation model for planning physical distribution systems: formulation of the computer model.

This dissertation deals with the computerization of a dynamic simulation model for planning physical distribution systems. Basic components of physical distribution systems, which include transportation, inventory carrying and handling, facility commitment and communications, are modeled to aid in the solution of long-range planning problems. Main topical areas include selection of the predominant computer programming language, description of computer subprograms in the model's supporting data, operating and report generator systems, and the model's operating system data base.

ALVIN MARTIN, Ph.D. Northwestern 1971. The decision-relevance of reported accounting information: an empirical perspective.

The study is an examination of the decision-relevance of accounting information (in annual reports) reported to common stock shareholders. We used a questionnaire to selected Chartered Financial Analysts which provided information as to the perception of the decision-relevance of particular annual report accounting variables. A multiple regression model was used to test the accounting variables respondents indicated were useful. The consistency of our statistical results supported the utility of the accounting data.

STEPHEN MEYERS, Ph.D. Pennsylvania 1970. A factor analysis approach to studying the structure of the stock market.

DIANE J. S. MIRACLE, Ph.D. Stanford 1971. The role of the egg futures market in the egg economy: 1940 to 1966.

Although focused on a futures market which ended in 1966, this analysis concerns controversies not yet fully resolved and relevant to all futures markets. The study suggests that hedgers are often better than the averag speculator in forecasting price trends in futures markets—the markets in which they are supposed to tak only an inadvertent and passive role in determinin prices. Analysis of the egg futures market in this perio indicates that hedgers made by far the greatest tota profits, contradicting a major premise of the Theory o Normal Backwardation (Keynes).

FLOYD E. MOAN, Ph.D. Michigan 1970. Managing in computer environment (an in-depth study of management of the inventory control function in five selecte companies).

ROBERT M. MONCZKA, Ph.D. Michigan State 197 Time-shared information systems for purchasing.

The purpose of this research was to provide new information on the use of time-shared information systems for purchasing and materials management. Fiel research was conducted with eight firms using o planning time-shared information systems and wit seven commercial time-sharing firms. Findings relate to the following were presented: 1) system application 2) design and development, 3) major changes, 4) computer systems, 5) future applications, and 6) commercia time-sharing.

JAMES E. MOON, Ph.D. Alabama 1970. An inquiry int the objectives and implementation methods of accounting computer education.

WILLIAM J. MORRIS, JR., Ph.D. Michigan State 197 Accounting for common stocks for church pensio funds: an empirical evaluation.

The subject is the valuation of equity securities hel as a long-term investment. The specific area studie was a determination of the valuation method for common stocks owned by church pension funds that bes meet their objectives for investment. The research included: development of specific criteria an accountin method should possess; development of a mathematica model incorporating the specified criteria to rate alternative accounting methods; application of twenty-on alternative accounting methods to common stock portfolios; selection and ranking of six acceptable method Both the Historical Cost and the Current Market Valu methods turned out to be unacceptable.

WAYNE J. MORSE, Ph.D. Michigan State 1971. Th allocation of production costs with the use of learnin curves.

Unit production costs frequently decline as tota production increases. A graphic representation of thi decline is called a learning curve. This research developed and tested a cost allocation model based on th learning curve phenomenon. Considerable attention wa given to the accounting concepts of matching an materiality. Solutions were proposed for a number o problems which may arise after the model is implemented. It was concluded that the model could be valuable tool in reconciling differences between th accounting period and the production life cycle of a product.

KENNETH S. MOST, Ph.D. Florida 1970. The role of accounting in the economic development of the modern state.

Economic theories are based upon an image of a preindustrial society; this does not affect the validity of the conclusions drawn by price theorists, but attempts to apply price theory models to problems of welfare and economic development founder on the inadequacy of the model of the production sector. One alternative is the accounting model, which admits time and space as explicit variables. The model is described and applied to the firm, the government agency, and the nation. French experiences using a similar model are described and evaluated.

EDWARD A. MOSES, Ph.D. Georgia 1971. Seasonal variations in common stock prices.

This study is an analysis of seasonal variations in monthly closing prices of common stocks listed on the New York Stock Exchange over the period July 1962 through March 1969. In addition to examining seasonality in individual common stock prices, an analysis is made of seasonality in market averages and of individual stocks in selected industries.

GEORGE J. MURPHY, Ph.D. Michigan State 1970. The evolution of selected annual corporate financial reporting practices in Canada: 1900–1970.

Selected practices are examined to reveal the various influences that have shaped corporate reporting practices in Canada over the years. The influence of America, England, Canadian corporate and tax legislation, the accounting profession and its various institutes, and of financial scandals is portrayed. Practices examined include the evolution of the mandatory audit and the content of the auditor's standard report, of the balance sheet and the income statement, and lastly of corporate depreciation practices.

CHARLES E. MYLER, JR., Ph.D. Florida 1970. Marketing policies of Florida's installment land sales companies.

PHILIPPE A. NAERT, Ph.D. Cornell 1970. Mathematical models of vertical market structures.

LARRY L. NELSON, Ph.D. Purdue 1971. Implementation of large scale planning systems: a study of computational dynamics and managerial interface.

This study was addressed to two major problems in the implementation of large scale planning systems, namely: computational dynamics and managerial systems interface. Efforts to increase computer processing efficiency included use of matrix computation techniques to improve the mathematical structure of the planning systems and use of automated matrix generator/report writer systems. An optimizing simulator was designed to enhance managerial systems analyst interface. It was used experimentally to study the decision-making and interface process within the context of tactical and strategic planning.

FRED NORDHAUSER, Ph.D. Purdue 1971. Foundational aspects of quantitative personnel selection research in the United States Air Force.

JAMES K. OWENS, D.B.A. Harvard 1971. A study of the uses to management of depreciation policy and practice.

The study is directed toward the role of depreciation policy and practice in the financial management of the nonfinancial corporation. The various definitions of depreciation, advanced both by external agencies and internal management, are carefully examined to ascertain points of similarity and contrast. The comparison yields some useful insights into present shortcomings and the possibility of alterations that might improve current depreciation reporting policy.

PETER G. K. PELLATT, Ph.D. California (Berkeley) 1970. A normative approach to the analysis of real estate investment opportunities under uncertainty and the management of real estate investment portfolios.

ABRAHAM PIZAM, Ph.D. Cornell 1971. Some socio-psychological correlates of innovation within industrial suggestion systems.

R. BURR PORTER, Ph.D. Purdue 1971. Application of stochastic dominance principles to the problem of asset selection under risk.

This thesis examines and extends the recently developed stochastic dominance procedures for analyzing the individual portfolio problem and compares these results with the well-known mean-variance and Theta models. It presents and demonstrates a new method for deriving comparative statics results concerning the equilibrium composition of portfolios, adds insights into the theory of innovative activity, and produces some further results concerning the concept of risk classes. A substantial mathematical appendix provides additional results.

VELAGAPUDI K. PRASAD, Ph.D. Michigan State 1970. An analysis of the purchase clustering patterns of food shoppers.

The purpose of the research was to investigate if shoppers who exhibited different degrees of food purchase clustering could be identified in terms of selected characteristics of shoppers. The characteristics examined were: 1) socioeconomic and demographic variables, 2) selected food purchasing characteristics of the shoppers, and 3) role-related self-perceptions of the homemakers. The research also examined if shoppers' patterns of purchase clustering among the food stores patronized are significantly related to the perceptual similarity of these stores.

JOHN H. PRONSKY, D.B.A. Harvard 1971. Pay system motivation: a comparative study.

The study compared the performance motivation of three pay systems, differentiated by the amount of pay information given to the managerial personnel. It was found that for a pay system to be effective, it must properly "fit," or is contingent upon, the organizational variables of task and people. A contingency matrix was devised to illustrate the variable relationships.

FRANK E. RANELLI, Ph.D. Alabama 1971. The objectives of large corporations: some empirical evidence.

FRED D. REYNOLDS, Ph.D. Alabama 1970. An analysis of the adoption process of retail merchants for a bank charge account plan.

RICHARD D. RIPPE, Ph.D. Harvard 1971. An econometric model of the income statement of the steel industry.

MARY LOU ROBERTS, Ph.D. Michigan 1971. The effect of information on consumer perceptions of and preferences for a grocery product.

JACK C. ROBERTSON, Ph.D. North Carolina 1970. A theoretical and structural design for independent audits of management representations.

This study makes a clear distinction between the concepts of audits for management and audits of management. The audits of management concept is examined in terms of managements' nonfinancial statement representations that are transmitted to the public. The research problems were to define, clarify and explain the term "management representations," and to determine whether an auditor could indeed verify them. These tasks were accomplished by utilizing the research methodology of journalistic content analysis.

CHARLES H. ROUSH, D.B.A. Harvard 1971. A study of the relationship between depreciation accounting methods and financial policy decisions.

The study examines how companies that use different depreciation accounting methods differ in terms of overall financial policies. Policies involving basic capital structure and cash flow decisions are examined by comparing sample data from companies using accelerated depreciation methods for both financial reporting and tax purposes with similar data from companies that use a straight-line method for financial reporting purposes and an accelerated method for tax purposes. The study shows that companies using different depreciation methods tend to be very different types of companies in terms of basic financial strategy.

JAMES M. ROVELSTAD, Ph.D. Michigan 1970. An investigation into the role of buyer-initiated, formal communication in marketing R&D to the government.

C. STEVENSON ROWLEY, Ph.D. Wisconsin (Madison) 1970. Inventory pricing in the grain industry: a study of current practice.

The study demonstrates that inventory pricing practice in the grain industry is consistent with generally accepted accounting principles, a proposition contrary to common belief. This contrary opinion is shown to result from failure to understand the financial reporting effect of hedging. Reasons for differences in reporting procedures between the grain industry and other industries, and among companies in the grain industry, are examined in detail, and reconciled. Suggestions for improved financial reporting are made.

PIERRE J. ROYER, Ph.D. Michigan 1970. Long-term warrants as financing instruments.

WALTER K. RUTLEDGE, Ph.D. Alabama 1970. Objectives of and criteria for interim financial disclosure.

WILLIAM L. SARTORIS, Ph.D. Purdue 1970. Analysis and administration of convenience and advantage licensing in the small-loan industry.

RICHARD F. SAUTER, D.B.A. Michigan State 1970. Demand estimation in a public utility: an empirical model.

This study provides increased understanding of the major factors that influence the demand behavior of customers in the industrial market segment of a telephone utility. It does so by (a) providing a conceptual framework for analyzing the variables that might affect the industrial concern's demand behavior, (b) employing empirical data to measure and test the degree to which these factors influence the demand for specific communications offerings, and (c) developing statistical models that specify the relationship between the most critical explanatory factors and selected demand criteria.

GARY G. SCHLARBAUM, Ph.D. Pennsylvania 1971. The investment performance of stock property-liability insurance companies: 1948–1967.

This dissertation is an empirical study of the portfolio performance of the property-liability companies. Risk-adjusted measures of performance are obtained for the companies which comprise the sample. These performance measures indicate that the company portfolios have performed as well as randomly selected portfolios. A decomposition of the overall performance measures indicates that the company portfolio managers possess no particular forecasting ability. However, investment expenses have been held to a minimum.

DAVID F. SCOTT, JR., Ph.D. Florida 1970. An inquiry into the existence of optimum financial structures.

This study tests the hypothesis that an appropriate range of financial leverage exists for firms in a given industry group. The traditional position is supported by the results of the one-way analysis of variance performed on the common equity ratios of seventy-seven corporations chosen randomly from twelve industries. The importance of earnings variability as opposed to industry classes on financial structures was resolved in favor of the industry surrogate for business risk.

WILLIAM B. SEALE, Ph.D. North Carolina 1970. Certain problems of shopping center development.

VERNON C. SEGUIN, D.B.A. Michigan State 1971. An investigation of the factors inhibiting growth of containerization in domestic surface freight shipments.

A questionnaire was used to develop a body of opinion concerning twenty factors found in current literature which relate to slower growth of van sized container shipments in domestic movements, as compared to foreign shipments. Freight users, carriers, and containerization equipment suppliers all felt that economic considerations, such as rate structures and empty container traffic, were the most important factors. Freight users showed concern over availability of equipment. Influence of government was not considered important.

TAYEL E. SHEHATA, D.B.A. George Washington 1970. Management development in the United Arab Republic: a comparative case study with special reference to the National Institute of Management Development.

The primary objective is to determine "What factors influence the effectiveness of management development institutes in the *UAR* as well as in developing countries in general." Data were obtained through a questionnaire and personal interviews, supplemented by secondary information. Findings form the basis for the recommendations for course to follow in the development of managers needed in the *UAR* with possible application for other developing countries that are undergoing industrialization.

SHMUEL D. SHRAIER, Ph.D. Columbia 1970. Consumer's buying behavior and the structure of retail industries.

Using Becker's theory of allocation of time and incorporating cost of holding inventory, a model of household buying behavior—time spent per trip and size of purchase—is developed. The structure of retail industries, i.e., the amount of resources employed and its allocation to stores, is shown to be affected by household time considerations. Support for this idea is obtained from empirical demand functions for the number of stores and persons engaged.

RICHARD H. SHUFORD, JR., D.B.A. George Washington 1971. The neglected problem in inventory budget decision making.

This study is concerned almost exclusively with the practical problems of inventory budget decision making. Economic order quantity (*EOQ*) model results provide the starting point for this analysis. The study seeks to answer the research question: "What effects do imposed budgetary controls have on the practical application of inventory management theory?" General propositions are developed as rational methodologies for suballocating budget dollars in organizations with multi-item inventories.

RAYMOND J. SHUSTER, Ph.D. Michigan State 1971. An exploratory study of the ethical perceptions of middle and lower level managers and the influence of nontechnical variables on their decisions.

The personal, influential, and environmental characteristics of 101 managers were examined for variables affecting their ethical perceptions. A test was made logically analyzing the ethical content of problem situations and rationally comparing decision choices of various managerial roles. It was found that organizational orientation influences the manager's ethical decision. Factors include: functional bias, differences in technical comprehension, variances in equating prudence and justice, and differences in importance ascribed to influences perceived as operative.

ANDREW F. SIKULA, Ph.D. Michigan State 1970. A study of the values and value systems of college roommates in conflict and nonconflict situations, and an investigation to determine whether roommate conflict can be attributed to differing values and value systems.

This study dealt with the problems of interpersonal conflict. The general methodology involved screen-interviewing some 1,200 college students in order to obtain a final sample of 100 roommate pairs (50 conflict pairs and 50 nonconflict pairs). The study found that, in general, conflict (or its absence) could significantly be attributed to differing (similar) values and value systems between college roommates. The study represents a significantly new and promising, as well as a heretofore unexamined and unresearched, approach to the study of interpersonal conflict.

PHILIP V. SPEARS, Ph.D. Kentucky 1970. A method for estimation of program economic contributions to college and university operations.

A model develops output-input ratios for administrators. Traditional human capital output is rejected. Program outputs are identified: education, research, public service, support, and miscellaneous. Inputs from program aligned ACE fund accounting system taken at current costs. Equivalent balance sheets, income statements, and planning schedules are developed. Steps in model development are detailed, and applied to Berea College (Kentucky).

STANLEY R. STANSELL, Ph.D. Georgia 1971. An examination of the relationship of reported research and development expenditures to common stock price-earnings ratio.

The study first examines the relationship of reported research and development expenditures to reported net income. A model is developed relating net income to four independent variables: total assets, number of employees, capital expenditures, and research and development expenditures using predetermined lags for the independent variables. Common stock price-earnings ratios are then regressed on the measure of R&D's effect on net income developed in the preliminary analysis in an attempt to determine whether *P/E* ratios reflect this relationship. The results indicate that *P/E* ratios reflect the full effect of R&D expenditures on net income.

WILLIAM T. STEVENS, Ph.D. Florida 1970. The development of multidimensional concept of management information systems, with particular reference to the conceptual framework of management accounting.

DAVID C. STUMBAUGH, Ph.D. Michigan 1970. An investigation into certain economic aspects of integrated logistics support planning for U.S. Navy weapon systems.

ALDEN P. SULLIVAN, D.B.A. George Washington 1970. An analysis of management training program evaluation practices in American industry.

Determining whether management training produces a favorable return on investment has long been an open question. The concept of training being an investment in human capital and handled as investment rather than period expenditure opens new vistas for studies in management. Progress toward ultimate answers in this area requires basic understandings of what is possible and of the state-of-the-art concerning evaluation of management training. This study provides some basic training evaluation insights.

JAMES D. SUVER, D.B.A. Harvard 1971. A comparison of the management control systems being utilized on commercial and government programs in the aerospace and defense industries.

This thesis compares the Management Control Systems being used to control cost schedule and technical performance on major government and commercial aircraft programs. Field research at eight major aerospace companies comprise the bulk of research data. The basic hypothesis that major differences do exist between the types of programs was subdivided into ten basic subhypotheses oriented to major subareas such as work authorization, schedule control, etc. Each subhypothesis was then tested and supporting data presented.

RONALD L. TATHAM, Ph.D. Alabama 1970. A taxonomic analysis of outdoor recreational participation patterns.

NORMAN E. TOY, D.B.A. Harvard 1971. Analysis of financial decisions for retirement: an exploration of several methodologies where there is uncertainty of investment returns and mortality.

This research, using as background university retirement programs with choice of fixed-income and equity securities, consists of four major parts. The first two deal with several subjects' use of interactive computer models of institutional, investment, and mortality environments prior to and after retirement. Subjects choose investment and actuarial strategies in view of uncertainties in investment returns and mortality. The third portion deals with assessment of utility for wealth at retirement based on previous simulated experience. Finally, utility structures and assessment procedures are developed for the retirement period explicitly considering uncertain mortality.

RONALD E. TURNER, Ph.D. Northwestern 1970. Perceptual dimensions of salesmen: a multidimensional analysis of call-allocating and sales-response behavior.

Techniques for obtaining subjective estimates of market characteristics from experienced salesmen were developed. The resulting measures were used to construct sales response functions for market analysis. A nonmetric multidimensional scaling analysis implied the number and types of customer attributes which were associated with sales call activity. A paradigm of call-allocating behavior by salesmen was proposed and was partially validated using multiple regression and discriminant analysis. Implications for marketing management were discussed.

RICHARD G. J. VANGERMEERSCH, Ph.D. Florida 1970. A study of institutional forces concerned with financial accounting in the United States, utilizing the annual reports of the U.S. Steel Corporation as reference points.

This dissertation attempts to isolate institutional forces affecting financial accounting, using the 1902–1969 annual reports of U.S. Steel as the reference point for discussion, and to draw conclusions about the effectiveness of each of these forces. These forces discussed in the dissertation are accounting professional bodies, federal government bodies, financial analysts and writers, stockholders, unions, and the New York Stock Exchange. Implications for today's financial accounting are drawn from this study.

JOSEPH G. VAN MATRE, Ph.D. Alabama 1970. Predicting success for initial public offerings of common stock: a discriminant analysis.

VEDANAND, Ph.D. Michigan State 1970. Role of product charisma in buying behavior: an analysis of black and white ownership of Cadillacs.

The study analyzes underlying consumer motivations and psycho-social differences of black and white buyers of Cadillacs. A theoretical model of charisma behavior was developed to explain the buying behavior of Cadillac owners; particularly those who belong to the subculture of alienation and poverty. Empirical testing of the data on attitudinal variables of model suggest the charismatic involvement of Cadillac buyers of the sample. It also suggests the existence of a market segment for luxury and prestigious products in the subculture.

HARISH L. VERMA, Ph.D. Michigan State 1970. Development of a joint-order inventory model.

The inventory control systems that are generally discussed in the literature are the fixed-order quantity system, the periodic review system and the (s, S) system where s is the reorder point and S is the maximum inventory level. These are independent ordering rules, i.e., the inventory model assumes that each set of item decisions is independent. Some joint-order inventory models are analyzed and another model is proposed. A simulation approach is used to study the economic advantages of the proposed joint-order inventory model over the "independent" reordering rules.

JAMES G. VERTREES, Ph.D. Michigan State 1970. Evaluation of alternative egg distribution channel systems.

This study attempted to answer such questions as: What factors affect costs among alternative distribution channel systems? What is the effect of administrative and exchange coordination upon distribution channel costs? What is the effect of instability in trading relations between firms upon channel costs? A comparison of channels—processing plant to distribution center to retail stores with plant direct to stores—indicated the average distribution costs of the two systems were similar.

Charles G. Walker, Ph.D. Alabama 1970. A residual equity theory of liabilities.

Orville C. Walker, Jr., Wisconsin (Madison) 1970. An experimental investigation of conflict and power in marketing channels.

This study viewed the distribution channels for consumer goods as social systems of interdependent organizations. It focused upon one specific type of interaction within these systems; conflictful behavior. A conceptual model was developed to explain the effects of different distributions of member dependence and power upon the forms and outcomes of conflictful behavior within channel systems. Propositions drawn from the model received empirical support from several industry case studies and a "mixed-motive game" laboratory experiment.

Richard G. Walter, Ph.D. Michigan State 1970. The effects of corporate financial policy on stockholder expectations.

A method for measuring stockholder expectations from the relative prices of a stock purchase warrant and its associated common stock was developed and applied to sixteen firms. Hypotheses regarding the effects of dividend policy and capital structure on the expected return were formulated and tested.

Donald E. Wiseman, Ph.D. Michigan 1971. The financial reporting of conditional ownership.

Frederick Wiseman, Ph.D. Cornell 1970. Market segmentation analyses on the buyers of new automobiles during the model year transition period.

Chorng-huey Wong, Ph.D. Minnesota 1970. Capital structure and investment behavior.

This thesis develops an economic theory which explains the behavior of the firms' financial structure and investment. Both the Modigliani-Miller approach and the expected utility maximization approach are considered. The two approaches are reconciled to some extent. An investment demand function is formulated in the context of expected utility maximization. The financial and the investment analyses are synthesized.

Thomas D. Wood, Ph.D. Florida 1971. A study of behavioral characteristics of conglomerates.

Eight behavioral characteristics were attributed to conglomerates. Forty-three conglomerates were used to accumulate data. Several implications for the future were included. There will likely be increases in: diversification among firms, the size and number of firms acquired, and acquisition price. Higher returns on investment will be required. The rate of merger failures will probably decline. Management efficiency and specialization will need more attention as well as a study of what motivates management.

Karl M. Zehms, Ph.D. Wisconsin (Madison) 1970. Municipal financial reporting.

This dissertation was concerned with three aspects of financial reporting: interim financial reporting, financial reporting to the public, and comprehensive annual financial reporting. For each of these three types of financial reporting, standards were set forth. The validity of the standards was established by discussing them with practicing municipal finance officers. Finally, actual municipal financial reporting practices of a sample of municipalities were evaluated in terms of the validated standards.

Robert A. Zelten, Ph.D. Pennsylvania 1970. Examinations of insurance companies.

The paper contains a theoretical and pragmatic discussion of the periodic, regulatory examinations of insurance companies. The study treats all aspects of examinations for solvency. Present examination practices are discussed and criticized. The present system is evaluated in terms of examination frequency, the scope of examinations and reporting thereon, examination costs and the personnel that conduct examinations. Reforms are offered to improve the present examination structure.

Industrial Organization and Public Policy; including Economics of Technological Change, and Industry Studies

Michael J. Baker, D.B.A. Harvard 1971. The diffusion of industrial innovations: an exploration of factors associated with the process, their measurement and predictive utility.

This thesis has the distinctive quality of bridging the economic literature on innovational activity and the business literature on marketing new industrial products. Through the use of regression analysis, using data from firms adopting two recent innovations, the significant attributes of customers most likely to adopt new products quickly are identified. It is then demonstrated how these attributes may be used by innovators to accelerate the diffusion process and, thereby, its profitability.

David J. Behling, Jr., Ph.D. Harvard 1971. Analysis of 1953–1975 changes in the structure of the *U.S.* petroleum refining industry.

A two-level general equilibrium model is developed which allows for variable input proportions and multiproduct industries. It is used to analyze changes in the input structure of the *U.S.* petroleum refining industry between 1953 and 1964 and to forecast the 1975 structure. Between 1953 and 1964 changes in product quality specifications, disembodied technological changes, and decreases in the relative cost of foreign crude oil were the most important factors explaining the increased capital intensity. Between 1969 and 1975, decreased tetraethyl lead consumption will lead to continued increases in capital intensity.

Louis K. Bragaw, Jr., D.B.A. George Washington 1970. Some characteristics of successful technological innovations, and their patterns.

Technological innovation is examined in intradisciplinary focus by building a model comprising phases of idea generation, problem solution and implementation. Organizational funding delineates the first two phases. Retrospective case studies of thirty-one new products or processes recognized as successful innovations by industrial research were drawn as samples to test the model by semistructured interview methodology. Data provided significant reinforcement for the model hypothesized.

Ben S. Branch, Ph.D. Michigan 1970. Research and development, profits, and sales growth.

In this study relations between research and development, profits, and sales growth are examined. R&D may increase or be increased by profits. Both of these possible relations are studied and compared in this dissertation. In addition, associations between R&D, sales growth, and its variance are considered. Data for a sample of 166 firms covering the 1950–64 period were used to test these relations. These firms were among the 1,000 largest manufacturers in 1950. For comparative purposes data from a second sample of 59 firms for the 1928–39 period was also examined.

Stewart E. Butler, Ph.D. Brown 1971. The effects of managerial attitudes toward risk on research and development project funding in small firms.

Three R&D project funding models are developed and tested against project expenditures of a case-study firm. The model which most accurately predicts the funding decisions of the firm employs a managerial utility function containing expected project profit and non-linear risk preferences. The model is shown to imply that risk-averse firms will invest less than firms with linear risk preferences, and that firms which are risk-preferers will invest more.

Rodger D. Carlson, Ph.D. Claremont School 1971. The economics of geothermal power in California.

Geothermal power plants will enjoy the advantage of having lower marginal costs than conventional plants as well as pollution free operation to preserve the environment. The advantages will be apparent for those areas which can utilize a geothermal power plant most fully because of a regular demand for power. Definitive analysis for a particular location would demand the consideration of other relevant factors such as capital costs and the size of the market area involved.

Sidney L. Carroll, Ph.D. Harvard 1971. An economic analysis of the airframe industry in the United States.

In this study the familiar framework of structure, behavior, and performance is used to gain some insights into the airframe industry in the United States during the period 1957–1966. After a rather careful definition of the industry, the structure in the government and commercial sectors is outlined separately. Next, price determination and other aspects of behavior are investigated and found to be markedly different in the two markets. Finally, profits performance in the separate areas is scrutinized.

Don R. Cook, Ph.D. Alabama 1970. Impacts of changing price levels on the monetary accounts of personal finance companies.

George E. Deal, D.B.A. George Washington 1970. The pricing of clinical laboratory tests.

This study attempts to obtain, on the basis of a national sampling of clinical laboratories in hospitals, in independent laboratories directed by pathologists or other doctors of medicine, or by physical scientists or technicians, the following information: the costing of tests and services; price policies and pricing factors for the tests and services provided by these laboratories.

Dennis P. De Melto, Ph.D. McGill 1970. The effect of foreign competition on the Canadian primary steel industry: 1950–1966.

This thesis analyzes how an increase in foreign competition has resulted in improvements in the operating efficiency of firms in the Canadian primary steel industry during the 1950–66 period. The performance of this industry is analyzed with the aid of a model based on recent developments in oligopoly theory as applied to situations where outside competitive pressures will force domestic steel firms to become more efficient by exerting a downward pressure on their prices and costs. The actual behavior of the steel firms during this period has conformed to the expectations generated by the model.

David A. Drinkwalter, Ph.D. Western Ontario 1971. Some economic characteristics of Canadian exporting firms.

This study examines the extent to which interfirm differences influence sales of Canadian manufacturing firms in foreign markets. Particular attention was paid to the amount of research conducted by the firms, the age of their productive facilities, their return on invested capitals, their asset base, and the age and educational level of their managerial bodies.

Isaac Ehrlich, Ph.D. Columbia 1970. Participation in illegitimate activities: an economic analysis.

A theory of participation in crime is developed and tested against data on crime variations across states and over time in the United States. The effect of probability and severity of punishment, unemployment, wealth and income inequality on the rate of specific crimes and the response of whites and nonwhites is predicted and estimated empirically. The empirical investigation deals with the interaction between crime and law enforcement activity and estimates the effectiveness of the latter.

Ali Ezzati, Ph.D. Indiana 1971. A computer simulation model for evaluating the dependency and bargaining power of the oil producing countries versus the international oil companies.

An empirical examination of the forces determining the relative bargaining strength of the oil countries and the companies. Parameter estimates are then used to build simulation models.

Muhammad O. Farruk, Ph.D. Cornell 1970. Structure and performance of the rice marketing system in East Pakistan.

A study to ascertain the extent to which there are imperfections in the working of the rice markets in East Pakistan and to diagnose the causes of any imperfections. The marketing system was found to have a large number of participants, likely to lead to a high degree of competition. The marketing system could expand relatively rapidly if large increases in agricultural production required such expansion but governmental action has largely been restictive of the private trade and would have to move towards more facilitative types of measures if modernization is to occur.

Edward L. Felton, D.B.A. Harvard 1971. The influences of new technology on the marketing of production inputs in a developing economy: the high yielding rice varieties in the Philippines.

The focus of this study is the revolution that has occurred in the marketing of production inputs for rice in the Philippines. It examines the changes that took place between the introduction of the high yielding rice varieties in this country in mid-1966 and the achievement by the Philippines of self-sufficiency in rice production a short two years later.

William P. Field, Jr., Ph.D. North Carolina 1971. An empirical study of the effect of market concentration on interest rates charged by commercial banks on small unsecured short-term business loans.

This dissertation dealt with the relationship between market concentration and market performance in local banking markets in the United States. Market performance was measured by interest rates charged on small, unsecured, short-term business loans, these interest rates being obtained by use of a questionnaire sent to each bank. The basic conclusion was that a marginal relationship between market concentration and interest rates charged seemed to exist, with higher concentration leading to slightly higher interest rates.

William J. Free, Ph.D. Missouri 1970. An interregional analysis of the soybean processing industry with particular emphasis on the southern region and the Tennessee Valley.

This study was a spatial equilibrium analysis with linear programming used to identify least-cost flows of soybeans, meal, and oil; and to evaluate the competitive position of the South under varying economic conditions. Different models were specified in order to investigate potential adjustments to changes in transportation and processing technology, rate innovations, as well as changes in regional supplies and demands.

Thomas G. Ganiatsos, Ph.D. California (Berkeley) 1971. Foreign-owned enterprise in Greek manufacturing.

David Gilbert, Ph.D. Harvard 1971. Mergers, diversification, and the theory of the firm.

Given the assumption of a well functioning capital market and profit maximizing management, two factors provide the basis for a negotiated acquisition: differences in expectations on the part of the acquiring management versus the existing management of the profitability of the potential acquired firm; synergistic potential. Using the alternate goals suggested by the so-called "new theories of the firm" (e.g., risk aversion and growth maximization), "manipulation" of the discount rate applied to the expected profit stream of the potential acquisition provides an additional basis for a negotiated acquisition.

Paul B. Ginsburg, Ph.D. Harvard 1971. Capital in nonprofit hospitals.

The problem of resource allocation in a private nonprofit sector is approached through a study of hospital investment. Based on a programming model, analysis of a sample of hospitals shows total investment to depend upon hospital size and capital availability while composition of investment is also dependent on demand and contributions for operations. Allocation is shown to be dominated by access to capital grants, a variable shown to be weakly correlated with an efficient allocation.

John R. Glennie, D.B.A. George Washington 1971. Public policy and the pharmaceutical industry: potential impact of proposed legislation.

This study's objective is determining effects on competition among pharmaceutical manufacturers if Congress were to enact certain legislation. The means of accomplishing this objective is testing hypotheses centered around three bills introduced into the 91st Congress. The passage of only one bill considered, the Montoya Medical Insurance Bill, was found to be potentially able to affect competition among pharmaceutical manufacturers. A major portion of the research deals with establishing how competition operates among pharmaceutical manufacturers.

Joel D. Goldhar, D.B.A. George Washington 1971. An exploratory study of technological innovation.

A study of the personal, organizational, and information flow variables associated with the innovative process. Recipients of the "IR-100" award were studied by self-administered questionnaire yielding a sample of 110 successful innovations. The data suggests that scientists and engineers are equally often innovators and exhibit the same type of satisficing behavior in information search as noninnovators. A new model for the "Design Concept Formulation" stage of technological innovation is suggested and a dichotomy is drawn between technology-means and economic-need-stimulated innovations, suggesting changing roles for technology transfer.

THOMAS W. HALL, Ph.D. Missouri 1971. Production and investment in the European steel industry.

Fixed proportion production functions, derived from engineering data, specify production process alternatives. Linear programming is used to estimate the optimal combination of steel producing activities in each ECSC country. The process analysis model is then modified into a dynamic framework incorporating investment. An econometric study of market demand is combined with a variation of the kinked-demand curve hypothesis to provide a rationale for stable market shares. The model is employed in an analysis of recent developments in the European steel industry.

TIMOTHY M. HAMMONDS, Ph.D. Cornell 1970. Utilization of protein ingredients in the *U.S.* food industry.

This study provides an economic analysis of protein rich food products which are available for use as protein supplements in fabrication of manufactured food products. The type and availability of protein sources are identified. The functionality of various protein-rich food additives as stabilizers, emulsifiers is identified. An attempt is made to project the type and extent of protein additive use which may be expected in the future.

RICHARD K. HART, Ph.D. Kansas State 1970. An investigation of discrimination in the rail rate structure with emphasis on regional incidence in the Missouri Basin area of the Great Plains.

JOSEPH M. JADLOW, Ph.D. Virginia 1970. The economic effects of the 1962 Drug Amendments.

The 1962 Kefauver-Harris Drug Amendments were designed to increase price competition for ethical drugs and enhance the safety of new drug testing. While successfully reducing the marketing of ineffective drugs, the amendments also have raised drug R&D costs and entry barriers to drug markets. These latter results have contributed to declines in the development of significant new drugs and in the competitiveness of ethical drug market structures.

RICHARD S. JOHNSTON, Ph.D. California (Berkeley) 1970. The growth of firms in some food marketing and processing industries.

It was postulated that data on size distributions of firms in some food and beverage industries were generated by a homogenous, first-order Markov process. Transition matrices were then estimated for these industries. A firm growth model was developed and the estimated matrices were used to test this model against Gibrat's Law of Proportionate Effect. The results were somewhat supportive of the growth model and cast doubt upon the predictive power of Gibrat's Law.

STANLEY M. KATZ, Ph.D. Pennsylvania 1970. A three-dimensional demand model of the competition between regulated rail and motor transportation in the United States.

This study seeks to create a transportation demand model which includes a competitive service variable and to test that model with empirical data for the purpose of establishing the relevancy and separability of price and service as explanatory variables in the demand for competitive transportation. A quantifiable form for the service variable is postulated, and data are collected from industry and government sources. Empirical results strongly confirm hypotheses of separability and relevance of these factors.

ALVIN KAYLOE, Ph.D. Colorado 1970. Resource allocation and control in the weapons acquisition process.

KEY WHAN KIM, Ph.D. California (Berkeley) 1971. The price and output of the Japan national railways, 1949–1963.

YUNG BONG KIM, Ph.D. Colorado 1971. The feasibility of establishing the sugar beet industry in Korea and some related policy problems.

JAMES T. KNEAFSEY, Ph.D. Ohio State 1970. An economic evaluation of mergers in the railroad industry—the C&O/B&O consolidation: a case study.

The study examines the impacts of large scale mergers in the railroad industry on various structural and performance characteristics. Among the more important hypotheses that are tested are those that relate merging to technical progress, size to technical progress, and pre-merger to post-merger technical progress. Two of the principal conclusions are that the costs of adjudication proceedings are excessive relative to the benefits and that, in the freight sector, mergers have been quite effective.

RUSSELL M. KNIGHT, D.B.A. Harvard 1971. A normative model for merger negotiations: the role of contingent contracts in the merger and acquisition decision.

The use of contingent merger contracts is reviewed and a normative model for merger negotiations proposed. Pareto optimal contingent contracts are derived. The study also describes problems and cures for contingent contracts and simulated merger negotiations where contingent contracts were introduced. These proved quite effective in facilitating agreement. The success or failure of a merger can often be traced to the price paid for the acquired firm. Contingent contracts helped determine a realistic price.

GEORGE F. KOPITS, Ph.D. Georgetown 1971. Dividend remittance behavior within the international firm: a theoretical and empirical analysis.

A hypothesis of foreign subsidiary dividend behavior is derived from maximization of worldwide profits by the international firm, constrained by internal financing of capital expansion. The hypothesis, estimated on published 1962 cross section data, performs satisfactorily for manufacturing *U.S.*-controlled foreign corporations established in developed host countries. Specifically, the findings suggest that taxation is an effective policy tool in influencing the dividend decision of overseas subsidiaries toward their domestic parents.

Joe F. Kososki, D.B.A. Harvard 1971. A model of airline fleet planning.

This thesis centers on an economic model for airline fleet planning. The model is designed to aid the fleet planners with their flight equipment purchase and sell decisions. The output of the model gives the best long-term fleet equipment decision consistent with the decision maker's feelings toward risk taking and his assessments of the uncertain environment, including the behavior of his competition.

Yehuda Kotowitz, Ph.D. Chicago 1970. Production functions in Canadian manufacturing 1926–61.

Philip A. Larson, Ph.D. Pennsylvania 1971. Group practice prepayment plans with special attention to rate making.

Robert J. Latham, Ph.D. Iowa 1971. The cost of medical education: an application of input-output analysis.

James D. Likens, Ph.D. Minnesota 1970. The welfare costs of nonoptimal airport utilization: a case study of the Washington-Baltimore airports.

A general transportation model is developed and applied to the three Washington-Baltimore airports. Delay functions are estimated for each airport, and estimates are made of input costs, including the value of travel time, to convert the delay functions into cost functions. Simultaneous equations are solved to estimate equilibrium marginal-cost prices and tolls for the system. The welfare cost of present nonoptimal utilization is $20 million per year, 20 percent of present trip costs.

Thomas L. Lyon, Jr., Ph.D. Missouri 1970. The rehabilitation industry: a health economic study.

This health economic study of the "rehabilitation industry" was designed to examine its structure, conduct, and performance. The rehabilitation industry was defined to include the manufacturers and fitters of rehabilitation supplies—orthotic and prosthetic appliances. Special attention was focused on the role of the industry's trade association in shaping pricing policies, quality standards, and general competitive attitudes among firms in this paramedical field. Discussion of bid pricing and distributions constitutes the major analytical work.

John M. McCarty, Ph.D. Southern California 1970. Economic aspects in the evolution of the Great Lakes freighter.

Since antiquity the contrivances for carrying commodities on the Great Lakes have been unique on the waters of the world. This study describes the technological characteristics possessed by commercial craft on the Great Lakes and uncovers and examines the economic factors that have caused the vessels on the Lakes to differ from their sisters that have contemporaneously worked the western rivers and sailed salt water.

Donald R. Marion, Ph.D. Massachusetts 1971. Economic analysis of food retailing in low income areas.

Daniel P. Maxfield, D.B.A. George Washington 1971. An appraisal of long-range planning for commercial air transportation.

The research is an analysis and an appraisal of current practice and performance of selected federal governmental agencies measured against a standard for long-range planning in support of the aviation industry. The appraisal identifies weaknesses in both the program areas included and the time span. A planning standard is developed as a descriptive model for commercial air transportation. Information was collected by survey of aviation interests and federal government officials.

Charles G. Moore, Ph.D. Northwestern 1969. An econometric evaluation of state electric utility regulation using a measure of manhours of regulatory input.

The objective is measurement of the "social value or return" of regulatory expenditures. Data consist of four cross sections on forty-six firms in twenty-six states for years 1947, 1953, 1960, and 1966. Results are statistically significant and indicate that regulation has had measurable perverse price and resource allocation effects but no effect on earned rate of return. A possible explanation is that regulators may have tried to cross-subsidize inelastic electricity users, thereby raising average unit costs in this decreasing cost industry.

R. Allen Moran, Jr., Ph.D. Massachusetts 1971. A price theory analysis of the role of cooperative advertising in the American economy.

Everett L. Morris, Jr., D.B.A. George Washington 1971. Earnings of the electric utility industry, 1966–68 based on regulatory accounting principles prescribed by the Federal Power Commission compared with what would have been reported if utilities had utilized generally accepted accounting principles prescribed by the accounting principles board of the American Institute of Certified Public Accountants.

Five specific subjects were studied to compare earnings of the electric industry based on regulatory accounting principles and generally accepted accounting principles. There were differences in returns and rates of return on rate base and to common equity. Regulatory accounting principles give utilities more flexibility to manage income. It was recommended that the Federal Power Commission amend its Uniform System of Accounts to make regulatory accounting principles more consistent with generally accepted accounting principles.

BRIAN M. NEUBERGER, D.B.A. Indiana 1970. Inducing socially desirable mergers: the potential role of minority interests protection.

In broad outline form, the dissertation considers first the problem of developing criteria for distinguishing socially desirable mergers from undesirable mergers. The law, suggested policy quidelines, and case examples of "good" and "bad" mergers are discussed. Next, the major hypothesis of the study is developed: protection of minority shareholders' interests can help insure socially desirable mergers. The basic definitions and issues are first considered, the current law on minority interests protection is described, and finally some proposals are made for strengthening minority positions.

PAUL H. NOWILL, Ph.D. Massachusetts 1971. Productivity and technological change in electric power generating plants.

BRUCE M. OWEN, Ph.D. Stanford 1970. The price effects of mergers.

A theoretical derivation of sufficient conditions for an increase in numbers of competitors to result in price decreases, and corresponding price effects for other kinds of firm interdependence, is followed by empirical analysis of the effect of joint ownership on prices in these industries: daily newspapers, television stations, electric power companies, and gas utilities.

JOHN RAPOPORT, Ph.D. Pennsylvania 1970. Product innovation: time, cost, and time-cost trade offs.

Empirical study of the time and cost required for various activities leading to product innovation, from applied research through manufacturing startup. Data from a sample of thirty-five recent industrial product innovations. The time-cost trade off function, a relationship between expected cost and expected lead time for an innovation project, is fitted statistically to estimates of time and cost made by project managers for each innovation. Analysis of the role of R&D in product innovation.

ROBERT J. REYNOLDS, Ph.D. Northwestern 1970. The regulatory process: an analysis of its impact on firm behavior.

A dynamic model of the regulatory process is developed and the optimal regulatory behavior in disequilibrium determined. Optimal firm behavior given this regulatory pattern is then analyzed. The theoretical implications of this theory are tested for the postwar period in the electric utility industry. The results are consistent with effective regulation and inconsistent with a number of alternative explanations.

PETER ROB, Ph.D. Arkansas 1970. The detection of price conspiracy in the electrical machinery industry through the use of economic data: a case study of a product line.

This study tests the hypothesis that the presence of a price conspiracy may be detected through the use of economic data. Specifically, it is suggested that during periods of active competition, the relationships between selected economic variables are consistent with the operation of the market as explained by existing economic theory. During periods of collusion these relationships are altered significantly and can no longer be explained by the normal functioning of market forces.

JAMES R. ROMAN, JR., D.B.A. George Washington. Opportunities for railroad passenger service in the northeast corridor.

The Penn Central Railroad has an opportunity to perform a valuable social service by attracting passengers to its service thereby alleviating congestion on the highways and in the air. The metroliner is already an attractive product, but the conventional train must be made more attractive before it is acceptable. Through proper pricing of both services, coupled with an aggressive promotional campaign, travelers should be willing to use these services instead of bus or air service.

ALLAN E. RUDOLF, Ph.D. Columbia 1970. A theoretical analysis of the Kibbutz as a producer's cooperative.

LOUISE B. RUSSELL, Ph.D. Harvard 1971. A cost model of Medicare.

A model of the decision-making process involved in choosing among alternative settings for medical care is developed starting from the premise that these decisions are primarily up to the doctor. Regression analysis is used to test the hypothesized relationships between state data on hospital utilization and costs under the Medicare program and various economic, demographic, and medical care variables. Medicare physician reimbursements and the effect of Medicare on the non-Medicare population are briefly investigated.

DAVID S. SALKEVER, Ph.D. Harvard 1971. Studies in the economics of hospital costs.

This thesis consists of four essays on the economics of hospital costs: A view of literature and hypotheses pertaining to hospital cost inflation; a study of the impact of product-mix change on hospital cost inflations; an econometric study of cost changes between 1961 and 1964 and between 1964 and 1967; an examination of the effect of uncertain demand on area-wide planning decisions regarding hospital size.

MOHAMMAD SAMI, Ph.D. Columbia 1971. The determinants of the *U.S.* demand for tourism.

FRANCISCO A. SANTOS, Ph.D. Columbia 1971. Factors affecting the shift of manufacturing industries to the southern region of the United States from 1954 to 1963.

Utilizing a measurement of industrial "market-orientation" this dissertation sets out to investigate whether the observed shift of manufacturing industries to the southern region of the United States had been systematically affected by the growth of the regional market. Such systematic effect was not found. The condition preventing such effect to obtain appears to have been—on the basis of an interregional comparison—the limited regional supply of skilled labor.

Lorraine H. Scheer, Ph.D. Oklahoma 1971. Influence of location factors in the Tulsa area plant location decision, 1958–1969.

This work initially justifies factors theoreticians conceive influence plant location decisions. Then, primary emphasis is placed on empirical analysis of the factors' influence on Tulsa area location decisions. The influence of the factors vary with product produced, employee size, and market area. This study further analyzes the factors' relative influence by major industry group, market area, and employee size. In general, the market situation and community environment appear as primary forces on location decisions.

Steven A. Seelig, Ph.D. Clark 1971. The impact of interest rate changes on price changes, by industry.

The hypothesis tested is whether interest rate increases lead to higher prices via a cost-push mechanism. Interest is viewed as a cost and its effect on supply is examined. A mark-up pricing model is developed and combined with equations relating interest rates to financial costs. This gives a set of equations which, using econometric techniques, enables estimation of the impact of interest rate changes on price changes by industry.

Edward R. Slack, D.B.A. George Washington 1970. The effect of the *U.S.* international telecommunications industry on the *U.S.* balance of payments.

The *U.S.* international telecommunications industry affects the balance of payments in the investment and the services accounts. For submarine ocean cables investment, the United States suffered a deficit of $50 million from 1956–1970. For satellite investment, a surplus of $115 million will have been received from 1966–1972. The $30 million deficit in the services account in 1967 could increase to $475 million by 1982. A to-be-developed telecommunications policy could change this situation.

Robert D. Smith, Jr., Ph.D. Massachusetts 1970. Cost problems of the Rutland Railroad and its successors (1937–68).

Paul E. Snoonian, Ph.D. Michigan State 1971. A comparison of industrial efficiency for Mexico, Puerto Rico, and the United States.

Mexican and Puerto Rican industries operating with *U.S.* levels of capital intensity do not achieve the same level of labor productivity. The resulting productivity gap is an "efficiency difference." Multiple regressions show high positive relationships between capital intensity changes and labor productivity changes. A Cobb-Douglas function is used to calculate efficiency differences among the compared countries.

Thomas R. Stauffer, Ph.D. Harvard 1971. Measurement of corporate rates of return on capital.

Stephen D. Stoller, Ph.D. California (Berkeley) 1971. The determinants of concentration change in American manufacturing industries: 1954–1963.

Rodney B. Swanson, Ph.D. Oregon 1971. Administered prices: an analysis of the concept.

John F. Trainor, Ph.D. Washington State 1970. The nursing home industry in Washington State.

Aged and infirm persons in the state of Washington and throughout the nation have turned increasingly to nursing homes. As these facilities are provided chiefly by proprietary firms their economic aspects are revealed by studying the operators as an industrial organization. Washington State has been one of the leading areas of the country in the development of nursing home care. The depth interview has been the fundamental tool or research in this study. Information obtained from the interviews was supplemented by other source materials for the purpose of corroboration and qualification.

Stephen L. Tucker, D.B.A. George Washington 1970. A comparison of major teaching and other community hospitals in terms of characteristics and activities of organization and administration.

Dynamic and progressive characteristics and activities of hospital organization and administration were determined using expert opinion. A comparison of questionnaire results received from major teaching hospitals and other community hospitals showed that in terms of their possession of these characteristics and activities there were statistically significant different rates of participation in a minority of items, with major teaching hospitals more actively involved in a large majority of the differentiated activities.

Jess E. VanDerwalker, Ph.D. North Carolina 1970. The supply and demand for publicly provided goods and services.

Philip Verleger, Ph.D. Massachusetts Institute of Technology 1971. A point to point model of the demand for air transportation.

Samuel Wagner, Ph.D. Pennsylvania 1971. An empirical study of the cost, time, outcome, and predictability of industrial research and development.

This study focuses on industrial research and development (R&D). The analysis includes the character of industrial R&D and its organization within the firm, the technical outcome of R&D projects, the use and market performance of successfully completed projects, the accuracy of estimated probability of technical completion, estimated cost, and estimated time in industrial and military projects, and the use and usefulness of R&D project estimates in project selection.

John C. Wassom, Ph.D. Indiana 1970. The relevance of the common carrier concept to the contemporary American transport system.

A brief outline of the historical evolution of the common carrier concept is provided. Some of the major problems that currently beset common carriers are examined. The responses of the regulatory agencies and the carriers themselves are provided. The costs and/or benefits of the current regulatory arrangements are examined. The possible impact of some suggested modifications to the existing scheme are examined. The conclusion is that the regulatory system we now have should be abandoned.

William D. Watson, Jr., Ph.D. Minnesota 1971. Costs of air pollution control in the coal-fired electric power industry.

Switching to low sulfur coal and precipitating fly ash from it is one way of reducing sulfur dioxide and fly ash discharges from a coal-fired power plant. This dissertation reports empirically determined cost and efficiency functions for such a process. Marginal cost curves are also developed and are used to determine the effluent charges required to induce a power plant operator to achieve given sulfur dioxide and fly ash removal levels.

John H. Wile, Ph.D. Brown 1971. The effect of technical change on the location distribution of an industry with application to the fertilizer industry.

Effects of changes in technical and demand parameters on long-run spacing of firms and long-run optimal price in market oriented industries were examined. Three models assuming, (i) constant marginal costs, (ii) nonconstant marginal costs, (iii) constant marginal costs and constrained output levels, were examined along with a generalized Weber framework. Each model was related to a fertilizer market. The hypotheses predicted by each were found to be, at least, not inconsistent with observed changes in these markets.

L. John Wilkerson, Ph.D. Cornell 1970. Factors influencing consumer acceptance of fluid milk substitutes.

Filled milk was introduced into the New York, California, and Arizona markets in 1967. This is a fluid milk product in which vegetable fat is substituted for butter fat. The study measures both the marketing experiences in these three areas and reports the consumer purchase response and attitudes toward the product by demographic classifications. The study covers a period of one year following product introduction.

Agriculture and Natural Resources

Mansoor Alturki, Ph.D. Colorado State 1971. Accelerating agricultural production in Saudi Arabia.

Resource allocation efficiency in Saudi's agricultural sector was analyzed by constructing two production functions for wheat and barley (Cobb-Douglas type). The value marginal productivities of land and labor were computed and allocation efficiency was investigated. Production from existing resources can be increased by removing disequilibria. Some barriers to efficiency were market structural variables, natural factors and government policies.

Mohamed H. Askalani, Ph.D. Iowa State 1970. Economic analysis and alternative procedures for estimating broiler production function.

This study covers three areas: 1) application of alternative procedures for estimating experimental broiler grain surface in terms of corn and soybean or energy and protein when faced with estimation problems related to autocorrelation and/or errors in variables; 2) specification of least-cost and least-times rations through marginal analysis and quadratic programming; 3) comparison between the outcomes of two broiler feeding experiments conducted in 1955 and 1968, respectively.

Azman Azis, Ph.D. Pennsylvania State 1970. A projection of Mexico's future requirements of base and precious metals and estimate of domestic potential supply.

Abdul Basit, Ph.D. Washington State 1971. Projections of demand and supply of wheat and rice in Pakistan, 1971–85.

Projections of demand and supply of wheat and rice were based upon analyses of relationships from 1959–60 to 1969–70 and upon assumptions regarding future population, income, and technological changes. Pakistan will likely become self-sufficient in wheat within the next three to five years, but only for about seven to eleven years. Surplus rice will likely be available each year until the mid-1980's when the country will again become a rice importer.

Sanford A. Belden, Ph.D. Purdue 1971. An analysis of decisions and performance in a production and marketing system for turkeys.

The implications of the interdependence of decisions at operating stages within a production and marketing system for turkeys were investigated. An intertemporal linear programming formulation was used to develop a complete optimization model of a firm comprising the grow-out, processing, and further-processing stages in the system. Specific emphasis was on decisions with respect to age (weight) and sex of turkeys, seasonality, facilities utilized, and inventories of products.

Richard A. Benson, Ph.D. Michigan State 1970. A comparative analysis of financing requirements of selected types of farm operations in the eastern corn belt for 1980.

This was a micro-demand study with implications for micro supply. Hypothetical farming units were constructed representing sizes and types of farming operations expected to be important by 1980. These included one, two, and larger than two-man operations for dairy, cash grain and feeder cattle. By 1980 such farm units will require more financing, more services, and more liberal credit terms than most farm lenders now offer.

James F. Berg, Ph.D. Ohio State 1970. The competitive position of the Ohio fed-cattle industry.

Richard C. Bishop, Ph.D. California (Berkeley) 1971. Factors affecting *U.S.* policy in ocean fisheries: a study in the political economy of resources management.

ALLEN C. BJERGO, Ph.D. Cornell 1970. A study of decision making in twenty-one New York farm families.

An examination of farm family decision making with a view to developing improved methods for extension teaching of decision-making principles; carried out as a phase of a study that was started a few years ago and is planned to extend for a total of fifteen years.

MICHAEL D. BOEHLJE, Ph.D. Purdue 1971. Strategies of creation and transfer of the farm firm estate.

The best strategy for estate transfer was ascertained for a case farm and selected variations from the case farm. The best strategy was "best" in the Bayesian sense, and the variable maximized was "size of estate transferred to the heirs." Mortality probabilities were explicitly considered. Substantial (taxable) gifts were part of the best strategies for all cases studied. The interrelationships between "creation" decisions and "transfer" decisions were found to be most important.

ROBERT W. BOHALL, Ph.D. North Carolina State 1971. Pricing performance of the marketing system for selected fresh winter vegetables.

Price analysis and the concept of site-price surfaces and functions were utilized to evaluate pricing performance. Pricing performance in the marketing systems for winter lettuce and carrots was generally consistent with performance models based on the competitive model in space, time, and form. In a limited number of areas, adjustments in institutions and market information systems may be needed. Results were less satisfactory for tomatoes reflecting the complexity of the form dimension of market price.

FRANK H. BOLLMAN, Ph.D. California (Berkeley) 1971. River basin development and the management of anadromous fisheries: an economic analysis of the Columbia River experience.

JOHN D. BRODIE, Ph.D. California (Berkeley) 1971. Transitional planning of forest management.

Alternative analyses of the forest resource base of Humboldt County, California, are provided utilizing simulation and linear progamming. Basic elements of the model are: a given stock-state; transition path goals for consumption of commodities and services and a planning horizon with a target stock condition. Eight levels of consumption curtailment are outlined. The results of random harvesting, and approximate "rule of thumb" criteria are compared with the linear programming optimum for one alternative.

RONNIE L. BURKE, Ph.D. Minnesota 1970. Technological and pecuniary marketing economies of size in Minnesota feedlots.

The major objective is to determine if marketing economies of size exist in marketing feedlot cattle in Minnesota, based on marketing resource alternatives and market alternatives. Consideration is given to short-run and long-run solutions and spatial factors related to feedlot size, feedlot location, types of markets, and market locations. Cost minimization with a linear program model is utilized to determine least cost solutions. It is concluded that marketing economies of size do exist related to more efficient use of fixed resources and pecuniary economies in procurement and dispersal of marketing resources.

KENNETH L. CASAVANT, Ph.D. Washington State 1971. An economic evaluation of the competitive position of Puget Sound ports versus Columbia River ports for Pacific Northwest wheat exports.

An important section of the Pacific Northwest wheat exporting channel is the regional port areas where the exporting activity takes place, Puget Sound and Columbia River. This study found wheat exporting activity will continue in both the Columbia River and Puget Sound ports although the Puget Sound ports will improve their competitive position with the Columbia River ports. Substantial increases in efficiency are available in the many technological and legal innovations considered in this study.

MUHAMMAD A. CHAUDHARY, Ph.D. Michigan State 1971. Performance of fertilizer distribution system in West Pakistan.

MEIR CHAYAT, Ph.D. Cornell 1971. Bargaining power to farmers and its welfare effects: a case study of the egg production industry.

This study outlines the types of bargaining prerogatives likely to be most useful to farmers and attempts to establish their consequences quantitatively. The egg industry, traditionally plagued by cyclical behavior, was chosen as the example. Although gains to society in the form of producers and consumer surpluses were estimated, it is likely that major benefits would accrue to the effects of stable prices upon investment in new technology and consequent shifts in the supply function.

JOHN H. CLEAVE, Ph.D. Stanford 1970. Labor in the development of African agriculture: the evidence from farm surveys.

This study attempts to determine how far African agriculture has developed from largely subsistence farming to participation in the market. It explores the problem of underemployment through a detailed analysis of seasonality in labor use and competing demands for labor time. There has been an impressive and widespread commitment of African farmers to production for market, largely for export; though new crops have been introduced, old hand methods continue in most areas; the level of annual labor hours in farming continues to be low, typically less than 1,000 hours, partly because of seasonality and partly because of opportunity cost; the flexibility and responsiveness of African agriculture is clear.

DONALD J. COCHEBA, Ph.D. Washington State 1971. The land market and the economic development process in San Juan County, Washington.

This is a case study of the land market of a county during a nearly complete conversion from an extractive industry economy to an economy based on retirement and seasonal residences and tourism. It explores the determinants of demand and supply in the land market and relates them to structural transition in the local economy. It also arrives at stages-of-growth hypotheses and forecasts.

MICHAEL H. COSGROVE, Ph.D. Ohio State 1971. Cost and quality of water service in Ohio cities.

CHARLES CUSKADEN, Ph.D. Michigan State 1971. Worker productivity in harvesting Michigan apples.

LEON E. DANIELSON, Ph.D. California (Berkeley) 1971. Investment in water resources with emphasis on timing.

This study develops an economic efficiency model for an analysis of optimal timing of investment in water resource projects. Investment decisions are considered as a decision process and solutions are generated using dynamic programming methods. Initial consideration is given to a staging model that determines optimal timing and staging simultaneously. A pure timing model based upon a fixed sequence of potential supply projects is then developed and serves as the analytical basis for this study. Empirical application was to the state of California.

THOMAS E. DAVES, Ph.D. Minnesota 1970. Economics of small watershed planning in Minnesota.

This study examines small watershed planning in Minnesota within the framework established by PL 566 and Minnesota water law. Institutions, analytical procedures, and the investment decision criterion are described and examined for their effects on economic efficiency. Proposed changes in the framework and procedures are tested empirically using data from thirteen PL 566 watersheds in Minnesota. Results indicate that most of the investments studied fail to achieve either economic efficiency or economic feasibility.

PAUL F. DAX, Ph.D. Columbia 1971. The plot-plantation system and the theory of land tenure.

BARRY W. ELLEDGE, Ph.D. North Carolina State 1970. Effects of government feed grain programs on beef and pork prices and costs to consumers.

This study investigated the effects of government feed grain programs on prices and production of both beef cattle and hogs and beef and pork. A mathematical model, used to predict changes in cattle and hog prices from alternative assumptions of changes in feed grain prices, showed that the loss in consumers' surplus was about $497 million annually for beef consumers and about $591 million annually for pork consumers. Consumer surplus loss, estimated by income groups, indicated that feed grain programs may have a "regressive" impact on consumers by income class.

JOAQUIM J. dC. ENGLER, Ph.D. Ohio State 1971. Alternative enterprise combinations under various price policies on wheat and cattle farms in southern Brazil.

DAN M. ETHERINGTON, Ph.D. Stanford 1971. An econometric analysis of smallholder tea production in Kenya.

Rapid increase in acreage of the smallholder sector of the Kenya tea industry makes predictions of future output important for optimal phasing of factory construction, repayments of loans to international agencies, and for quota allocations under the International Tea Agreement. This study concentrates on the specification and estimation of a multiperiod production function for smallholder tea. The study brings out the need to concentrate on labor-intensive technology—since tea is labor-intensive to an almost unique degree among agricultural crops—and to rely on the profit motive of farmers.

DON E. ETHRIDGE, Ph.D. North Carolina State 1970. An economic study of the effect of municipal sewer surcharges on industrial wastes.

A theory of the firm's response to sewer surcharges, a charge based on the quantity and strength of wastes discharged, was developed and the response estimated by two different procedures for two different industries: the beet processing and the poultry processing industries. A synthetic or budgeting approach was used to estimate the response for both poultry and beet processing firms; a regression analysis using both cross section and time-series data was used to estimate the response for poultry processing firms.

DAVID E. EURESTY, Ph.D. Pennsylvania State 1971. A systems approach to the regional evaluation of potential mineral resources using computer simulation, with a case study of the impact of infra-structure on potential supply of base and precious metals of Sonora, Mexico.

ROGER K. EYVINDSON, Ph.D. Iowa State 1970. A model of interregional competition in agriculture incorporating consuming regions, producing areas, farm size groups and land classes.

In order to study interregional competition in American agriculture, a linear programming model of this industry was set up. For this model the United States was divided into 21 consuming regions and 157 producing areas. The farms in each area were divided between three size groups. Production activities for the principal crops and livestock of *U.S.* agriculture were established for each size group, and production on each of these groups was limited by the available land, labor, and capital. The model also includes transportation activities for the products considered.

RICHARD F. FALLERT, Ph.D. Purdue 1971. An analysis of buyer-seller relations between food chains and fluid milk processors in the North Central region.

Food chains increasingly are negotiating terms of trade at their division, regional, or national offices and initiating various forms of backward vertical integration into milk processing. Sixty-five percent of the food chains and wholesaling organizations in the North Central Region used some form of centralized fluid milk procurement program at the time of the 1968–1969 survey. Program savings came chiefly from lower milk cost, economies obtained by dealing with fewer suppliers and reduced delivery services.

RONALD D. FINSTER, Ph.D. Arizona 1971. Water policy implications of a changing economic structure for Arizona.

Water policy in arid regions is almost exclusively supply-oriented. This study develops a demand-oriented water policy and attempts to ascertain whether it offers a viable alternative. Demand-oriented policy, which aims at decreasing the quantity of water demanded, is effected through changes in the productive structure and external trade. The efficacy of demand-oriented policy is tested by using an Arizona input-output model. Results demonstrate that such policy is quite effective.

BARRY L. FLINCHBAUGH, Ph.D. Purdue 1971. The 1968 voluntary feed grain program: characteristics of participation and intrafarm cropland diversion in Indiana.

This dissertation estimates the intrafarm difference in productivity between diverted and farmed acres and develops profiles of participants and nonparticipants in the 1968 Voluntary Feed Grain Program. A sample of 240 randomly selected corn farmers in four corn belt counties of central Indiana were visited and interviewed employing a questionnaire predesigned to collect facts, attributes, attitudes, and opinions from both participants and nonparticipants in the 1968 program. Diverted and farmed acreages of the participants were drawn on corn productivity coded, soil survey maps.

MICHELE B. FRANCOIS, Ph.D. Tulane 1971. Cost functions of the Louisiana sugar cane cooperatives.

JEREMIAH E. FRUIN, Ph.D. California (Berkeley) 1971. A linear programming model of a multiplant tomato packing firm.

A multilocation firm with limited capacity processes a seasonal commodity into many finished products. Raw product characteristics vary and affect yields and costs of finished products differently. Management desires to optimally assign each lot of commodity to the plant and finished product which will maximize profit subject to capacity and demand constraints. A model of an existing tomato processing firm with these characteristics was formulated and solved using the decomposition principle of linear programming.

STEPHEN W. FULLER, Ph.D. Kansas State 1970. The optimum number and size of country grain elevators in spatial equilibrium.

The primary objective of the study was to determine optimum number and size of country grain elevators in a spatial equilibrium framework. The results of this study allowed analysis of market performance, insofar as performance was determined by market structure. The model, employed by this study, minimized internal plant cost and assembly cost. When the existing market structure was compared with market structure as determined by the model, existing plants were too small and numerous.

TERRENCE F. GLOVER, Ph.D. Purdue 1971. Seasonal employment of hired labor in *U.S.* agriculture.

This study is concerned with analyzing the seasonal demand and supply for hired agricultural labor. A two-equation, three-seasonal model of seasonal demand and supply for hired agricultural labor was estimated. The model was estimated both on a national and regional basis. Seasonal differences in the elasticities of supply and demand regarding the wage rate were found. Strong links between seasonal hired farm labor market and general economic activity were indicated.

LYNDON E. GOODRIDGE, Ph.D. Purdue 1971. Evaluation of information for agricultural replacement decisions.

The expected value of perfect information for agricultural equipment investment/replacement decisions with respect to timing of firm growth; amount of firm growth; and level of machinery failure. The decision options included owning, renting, leasing, and the custom hiring of needed machinery capacity. Information with respect to timing and amount of firm growth was of considerably greater value than information about probable machinery failure. Also, the discounted net revenue function was sensitive to changes in interest rates.

LOIS E. GOSSE, Ph.D. Cornell 1970. The demand for outdoor recreation.

BROOK A. GREENE, Ph.D. Cornell 1971. Rate of adoption of new farm practices in the Central Plains, Thailand.

A study of diffusion of five production-increasing innovations in the Central Plains of Thailand. The work emphasizes the great variety of diffusion patterns as innovations of differing characteristics interact within varying environments. The complex interactions of economic and social factors are documented. Major conclusions support the concept that the absolute may be more important than the relative rate of return in encouraging new innovations and that factors other than rate of return seem to play a key role in adoption.

JUDITH M. GUERON, Ph.D. Harvard 1971. The economics of solid waste management.

The services of solid waste collection and disposal, presented as stages in a waste generation-absorption cycle, are defined, their public good qualities distinguished, and the available cost and demand information summarized. The sale of collection to individual consumers and disposal to local governments is evaluated. The interlocal exchange of pollution abatement services is examined using a public good "trade" model. A national "pollution added tax" is discussed that would force consumers and producers to consider the social costs of residual generation.

JOSEPH F. GUINAN, Ph.D. Iowa State 1971. Estimation of egg production functions from experimental data with economic applications.

Egg production functions are estimated using data from a poultry nutrition experiment. Ration ingredients are used as X-variables with particular emphasis on problems of estimation resulting from ad libitum feeding and multiple observations on the same experimental unit. Economic and biological implications of the estimated functions are discussed.

JOHN K. HANES, Ph.D. Minnesota 1970. Price analysis approach to market performance in the Red River Valley potato market.

This study presents a conceptual framework and analytical approach for objective measurement and evaluation of price performance in agricultural markets. A synthesis of three different theoretical price models is used to analyze "real world" prices in the context of various modifications and extensions of the perfect market concept. The method of analysis is specifically applied to an evaluation of price performance in the Red River Valley potato market.

GARY D. HARTER, Ph.D. Iowa State 1971. Estimates of a dairy cow milk production function.

Data for 165 dairy cow lactions were combined to estimate milk production from feed inputs of alfalfa hay and a concentrate mixture. Estimates of whole milk and butterfat production were combined to form value of product functions and to estimate butterfat percentages for various feed combinations. Computer iteration procedures were used to predict the ration combination for maximum profit for given prices. Estimates of ad libitum intake and the shape of the production surface were important in determining optimum feed combinations.

PETER B. R. HAZELL, Ph.D. Cornell 1970. Rational decision making and parametric linear programming models for combining farm enterprises under uncertainty.

A family of multiple objective farm planning models for enterprise combination under uncertainty was developed and tested. The portfolio analysis approach using quadratic programming was included with parametric linear programming models each involving expected income and one of several alternative measures of risk. Consideration for selecting and estimating these models led to the designation of the expected income and mean absolute deviation model as likely to be appropriate under most circumstances.

RICHARD J. HERDER, Ph.D. Minnesota 1970. An analysis of agricultural credit lending by commercial banks in the Ninth Federal Reserve District.

ROGER W. HEXEM, Ph.D. Iowa State 1971. Factors affecting the economic and social well-being of agriculturalists in less developed areas and agriculture's contribution to general economic expansion.

To systematize a discussion of the impact of economic and noneconomic factors on the economic and social well-being of agriculturalists in less developed countries, a micro-oriented model of agricultural growth is developed. Within this construct, some of the "friction points" of agricultural development are observable together with their qualitative impact on developmental processes. The implications for national planning to expand growth of the agricultural sector are discussed.

GERALD L. HORNER, Ph.D. Washington State 1971. An appraisal of alternative land uses in the Skagit County trade area of Washington.

A Leontief-type gross flows table was converted to a linear programming model of the area economy to estimate area economic impacts of alternative uses of river valley and delta land. Retention of primarily agricultural use of this land need not be costly in terms of foregone employment and gross regional product in the area. However, the maximizing criteria of gross regional product and employment yield different optimizing economic organizations.

JOHN E. HOSTETLER, Ph.D. Michigan State 1970. Sensitivity analysis of selected linear programming assumptions: a study of the stability of agricultural projections in river basin research.

This study was undertaken to evaluate the extent and direction of certain errors in assumptions used in river basin projection models. Sensitivity analysis was applied to five types of assumptions. Projected demand, livestock feeding relationships and adoption rate of new technology were relatively sensitive assumptions. Soil management practices and minimum acreage constraints were relatively insensitive in their effects.

JOHN H. HUMPHREY, Ph.D. California (Berkeley) 1971. Resource allocation and income distribution in agriculture: a case study of an irrigation economy in northwest Mexico and its implications for development.

This study focuses on the relationship between agriculture and the economy during development in Mexico since 1930. Emphasis is placed on the demand by the agricultural sector for industrial consumer goods. A relationship between changes in income distribution in agriculture and the size and pattern of such demand is estimated. A more equal distribution of income will bring about: a rise in total consumer demand; a fall in private savings; a shift in demand to products of more labor intensive industries; and a swing of demand toward domestically produced goods.

RALPH HUTCHINSON, Ph.D. California (Los Angeles) 1970. The California wine industries.

This study traces the early history of the California wine industry, outlines the events of the Prohibition era, and examines in detail post-Repeal resource allocation problems of the wine industry. The study demonstrates that Prohibition was directly responsible for two major problems faced by the California wine industry in the post-Repeal period: 1) securing appropriate quantities of grape varieties suited to wine production; 2) matching the composition of wine production to changing demand for wine.

JEROME E. JOHNSON, Ph.D. Minnesota 1970. Interrelationships between land tenure and progress in North Dakota farming.

This study examines how land tenure arrangements and tenure-related factors affect the economic progress of North Dakota farm operators. Economic progress was defined as total dollar change in net worth over a four-year period. The analysis used multiple regression to examine factors affecting changes in net worth, which was separated into four components. Data were gathered by interviews from a random sample of Red River Valley farm operators.

RUSSELL B. JOHNSON, Ph.D. Missouri 1970. Agricultural loan evaluation with discriminant analysis.

This study was designed to identify those factors which influence the financial success of selected agricultural borrowers and to develop a computerized credit scoring model for evaluating loan applications. Discriminant analysis was used to select the important financial variables and assign appropriate weights. The accuracy of the credit scoring model was then tested using loan data from samples of production credit loans.

EDWARD H. KREBS, Ph.D. Michigan State. Simulated price and supply control programs for the Michigan bean industry.

A navy bean industry supply and demand model was estimated via a combination of ordinary and three-stage least squares. The estimated supply-demand relationships were the basis of a recursive simulation model used to examine different private or government price and supply control programs. Such programs (especially government support programs) would have increased producer income from navy beans above what it would have been without controls.

ROBERT G. KULLER, Ph.D. Kansas 1970. Optimal exploitation of petroleum resources.

A model of an oil and/or gas producing firm operating under competitive conditions is presented. Production and cost functions incorporate physical laws governing the process. Assuming profit maximization over the economic life of the reservoir, a characterization of the optimal time path of production is developed. It is shown that a corporate income tax does not alter the optimal production path. To obtain meaningful statements concerning a depletion allowance, specification of functions is indicated.

STEVEN W. LAMB, Ph.D. Kansas State 1970. Estimating efficiencies in providing rural roads.

LAWRENCE W. LIBBY, Ph.D. Cornell 1970. The political economy of water management: conceptual model and decision strategy for the Susquehanna Basin.

Theory in economics and political science is reviewed, and based upon extensions and synthesis, a model for water resources planning is developed. This is applied to results of recent planning studies of the Susquehanna River Basin. Since an intergroup bargaining framework is the basis for the model, considerable reworking of the data is required. Major shortcomings of the planning process as it exists are identified and corrective measures suggested.

C. C. LIM, Ph.D. Missouri 1971. An econometric investigation of feed utilization for Missouri dairy farms: a static analysis.

Linear, exponential, quadratic, and square root types of production functions were estimated empirically using daily milk production data from DHIA records. Dummy variables and analysis of variance models were used to investigate herd size, breed, and feeding level effects. Empirical results indicated that breed and concentrate level effects were statistically significant whereas herd size effect and the effect of different combinations of succulent forage and dry forage were not significant.

EDNA T. LOEHMAN, Ph.D. Purdue 1970. Optimal resource allocation and some techniques of optimization.

This dissertation deals with optimal allocation of water resources and waste treatment when there are pollution externalities. The second best welfare relation between competitive equilibrium and Pareto optimality under a pollution externality is considered as well as the value of piecemeal policy. The effect of changing stream flow over time on optimal waste disposal in a stream and on optimal regional treatment plant design is considered. A zero-one integer programming technique is presented to determine optimal waste treatment levels, plant size, and location. The results of decentralized decision making by interested parties on financing and choice of a regional treatment system is considered.

MAX I. LOYD, Ph.D. North Carolina State 1970. Availability and consumption of dairy products in eastern North Carolina.

Survey data were analyzed using multiple regression to determine the effect of availability on consumption, with other socioeconomic variables considered. Consumption of fresh products increased as income and family size increased, and as shopping distance decreased. Consumption of substitute products was positively correlated with family size, food assistance and shopping distance; and negatively correlated with income. Home delivery had a positive, though generally nonsignificant, effect on the consumption of fluid milk products.

JACK J. MCCULLICK, Ph.D. Kansas State 1970. The economic impact of irrigation on selected southwestern Kansas counties: an input-output approach.

Using input-output analysis, the basic objective of the dissertation was to quantitatively measure the economic impact of irrigation on local communities in a four county area in southwestern Kansas. By applying output and income multipliers to the net increase in agricultural production that is attributable to irrigation, the impact on personal income and input sales transactions was quantitatively determined. Based on the multiplier analysis, some general strategies relating to economic development in the four county area were discussed.

BENNY R. MCMANUS, Ph.D. Purdue 1971. Credit reserves for midwestern farmers.

This study ascertains workable strategies for farm lenders and borrowers in adapting to risk and uncertainty for debt repayment. Specific objectives were to: estimate the income variability and attribute it to various sources in the cases of selected farms in central Indiana: estimate the amount of debt relative to real estate assets that selected farms could have serviced for the 1949–1968 periods; ascertain the influence of income variability on debt repayment capacity; and evaluate alternatives for providing financial reserves to cope with income variability and debt repayment.

KIGHOMA A. MALIMA, Ph.D. Princeton 1971. The economics of cotton production in Tanzania: an examination of some of the factors that influence agricultural development.

This study examines some of the factors which enabled cotton production in Tanzania to increase at 10.0 percent annually between 1953 and 1969. It has been undertaken in the hope that the results would shed light on some of the factors that generally influence agricultural development. Consequently, cotton farmers in Tanzania have been found responsive to the producer price. In addition, cotton research, infra-structure and marketing facilities played a complementary but somewhat crucial role.

MAHAR MANGAHAS, Ph.D. Chicago 1970. An economic analysis of the diffusion of new rice varieties in Central Luzon.

This is a study of determinants of the use of new high-yielding varieties of rice in Central Luzon in the Philippines. Linear probability functions were estimated from a cross section of 866 farmers for crop year 1967–68. In addition to irrigation and participation in the extension program, previous knowledge of modern cultural practices is found to be an important determinant of varietal acceptance. In contrast, general schooling, farm size, and farm ownership are of very minor relevance.

CHARLES K. MANN, Ph.D. Harvard 1971. Taste, technology and the government: the case of tobacco.

Taking tobacco as a case study this analysis focuses on interactions between government, technology, private industry, and consumer taste. Changes are predicted in the demand for flue-cured and burley tobaccos in the government tobacco programs and in the production technology. Several proposals are advanced to cushion the dislocation foreseen: graduating allotment reduction by income status and designing conversion programs to concentrate tobacco acreage reductions in areas with the most promising nontobacco alternatives.

DAVID MISHAELI, Ph.D. New Mexico 1971. A dynamic decision model evaluating alternative policies for long-run water allocation.

A long-run decision model for water allocation in agriculture and evaluation methods for alternative public policies regarding temporal water allocation, were developed. A set of decision rules and a computation technique (variational programming, utilizing calculus of variations) were developed as a basis for the dynamic decision model with a multigoal objective function. The economic signals to determine optimal public policies were obtained by comparing the trajectories generated by two different organizational models (mono-decision and poly-decision).

JAMES E. T. MONCUR, Ph.D. Washington State 1971. A programming approach to the valuation of water in alternative uses.

The question of water resource evaluation is approached in terms of value foregone due to a substantial extra-basin diversion. A planning model is constructed to determine optimal allocation of a given amount of water, along with associated "products"—firm and dump energy, irrigation and downstream water releases, and water quality—for each (monthly) time period and each reservoir or reach of the system. Optimal allocations and net revenues are determined in a model which adapts the decomposition principle for linear programs to the case where some constraints involve non-linearities.

WILLIAM D. MONTGOMERY, Ph.D. Harvard 1971. Market systems for the control of air pollution.

This thesis deals with the development of a licensing system for efficient control of air pollution caused by industrial sources at fixed locations. With the assumption that a desired air quality is given exogenously, the problem reduces to the standard one of allocating a scarce resource—waste disposal capacity of the atmosphere—among potential users. When rights to use this capacity are made transferable, private bargains among firms can lead to least cost achievement of the air quality standard. Under more stringent conditions, which do not include the absence of externalities, but do include general convexity of cost functions, a competitive market in licenses can produce this result.

KIRBY S. MOULTON, Ph.D. California (Berkeley) 1970. Public policy and market effectiveness in the Douglas Fir lumber industry.

JON L. NEVINS, Ph.D. North Carolina State 1970. A model for calculating the optimum analysis of mixed fertilizers.

A new procedure was developed for simultaneously calculating the optimum concentration of a mixed fertilizer (using linear programming and assuming a given ratio among nutrients) and the optimum size of the mixing facility. Conventional analyses tended to recommend products that were too highly concentrated. The new procedure did a better job of explaining the timing of recent increases in the concentration of North Carolina's fertilizers than did the earlier prescriptions.

GERALD M. NOLTE, Ph.D. Minnesota 1970. Effect of milk assembly and processing costs on optimal type and size of plant for butter and nonfat dry milk in Minnesota.

The economics of size in processing milk into butter and nonfat dry milk and the diseconomies of size in milk assembly are analyzed. The analysis includes an investigation of the size economies of the several major processing activities and the effect of carrying on these activities in separate plants. The analysis also included an investigation of the cost of direct farm-to-plant milk assembly and indirect milk assembly via local milk receiving stations.

JOSEPH K. OLAYEMI, Ph.D. McGill 1971. Peasant cocoa production in Western Nigeria: an economic analysis.

This thesis examines some economic aspects of the peasant cocoa industry in Western Nigeria: the characteristics of farm resources and farm organization; the expected rates of return from cocoa production; available technologies and their rate of adoption; farmer's supply responsiveness to price and other variables; and, the impact of Marketing Board's pricing policies on cocoa production and revenues. From the analysis, it is concluded that cocoa is the most profitable crop available in the agricultural region; that expected profit is only one of several considerations that influence the adoption of new technologies; that the farmers' response to price is inelastic; and that the Marketing Board's policies have depressed cocoa production and revenues.

JAMES L. PEARSON, Ph.D. North Carolina State 1970. A growth model for evaluating investment alternatives of a processor of seasonally produced products: an application to vegetable canning.

A multiperiod mixed-integer programming model was developed to identify production and investment decisions that maximize a firm's growth over a planning horizon. The model selects product mixes and line capacities, allocates capital, and invests and disinvests in equipment. Application of model was made to assess economic feasibility of investing in a vegetable canning plant. Effects of capital withdrawals, factor and product price changes, supplies of raw products, and demand for finished products on growth were evaluated.

NEIL H. PELSUE, Ph.D. Purdue 1971. An economic analysis of the impact of emerging Grade A milk supplies upon the North Central dairy industry.

Two forces are encouraging the conversion of farms producing Grade B milk to Grade A milk standards—price incentive and increasing number of manufactured products requiring Grade A milk. This study attempts to determine the nature and extent of the impact of Grade A milk supplies upon the Federal Milk Order pricing system. A simulated model of the North Central dairy industry is constructed. This study indicates continued declines in numbers of milk cows. The Grade B to Grade A conversion is expected to continue. Class I and II prices are expected to continue their upward trend.

THOMAS A. PETERSON, Ph.D. Washington State 1971. An analysis of thirteen years of commercial hail suppression in central Alberta.

Statistical analyses were conducted to test for reduced hail damage to crops during a thirteen-year hail-suppression program. Benefit-cost projections were also made of hail reduction in relation to suppression costs. Results suggested only minute reductions in hail loss are required for successful suppression programs. Because weather modification in Alberta has been a controversial subject, it was strongly recommended that steps be taken to determine legal, social, and economic implications of wide-scale suppression.

WILLIAM H. PIETSCH, Ph.D. Washington State 1971. An input-output analysis of import substitution: a case study of potential impacts of expanded Washington beef production.

Leontief-type input-output models in two forms (paired regional and interregional) were used to estimate the impacts of an import substitution strategy on the economics of both affected regions. Changes in total income as well as the distribution of impacts among sectors and between regions were considered. The analysis consisted of two case studies. Each study involved the substitution of Washington produced beef products for those currently imported.

CHARLES G. PLOURDE, Ph.D. Minnesota 1970. Optimal exploitation of a replenishable natural resource.

This dissertation considers the problem of optimally exploiting common-property replenishable natural resources using control theory. The first model assumes that humans reduce the biomass growth rate through consumption. Assuming costless production and positive time discounting it follows that the maximum sustained yield biomass exceeds the optimal steady state biomass. When production costs are introduced the optimal steady state biomass increases. Methods of governmental regulation to internalize the common-property externality are analyzed.

JOHN T. PORTER, Ph.D. Purdue 1971. Economic and marketing potential for expansion of horticultural production in the lower Wabash Valley.

The following crops appeared to offer the most promising production potentials: strawberries and tomatoes for fresh market; cantaloupes, watermelons, and potatoes; and snap beans for processing. Watermelons and cantaloupes were found to offer the most likely opportunities for acreage expansion. Potatoes will provide the greatest potential for increasing gross farm income. The lower Wabash Valley has a very favorable competitive position in truck transportation to nine major Midwest markets.

GEORGE R. PRIGGE, Ph.D. Ohio State 1970. A simulation study of tenure and size relationships of selected corn-hog farms.

FREDERICK J. PROCHASKA, Ph.D. North Carolina State 1970. Opportunity cost of time and other socioeconomic effects on away-from-home food consumption.

The effects of the opportunity cost of homemaker's time, household income, and other socioeconomic variables on away-from-home consumption were investigated. The value of homemaker's time was found to be significant and positively related to the number of meals consumed away-from-home. Family income was found to also have highly significant positive effects, but not as great as would have been erroneously estimated if differences in opportunity cost of time had not been considered.

RUSSELL RANEY, Ph.D. Ohio State 1970. An explanation of changes in the number of farms in the United States, 1910–64.

ROBERT A. RATHJEN, Ph.D. Minnesota 1970. An economic analysis of fertilizer retailing in Minnesota.

The general objective of the study was to estimate the probable changes in the economic structure of the Minnesota fertilizer industry by 1980. This was accomplished by analyzing: 1) expected demand for fertilizer; 2) cost functions for alternative types of firm organization; and 3) consumer preferences for products, services and types of firm organization.

WILLIAM N. ROSS, Ph.D. Kansas State 1970. Empirical relationship between tenure and economic growth of farms in central Kansas, 1940–1962.

Through the use of analysis of variance and multiple linear regression models the study associated selected measures of economic growth with the land tenure arrangements of 220 randomly selected commercial farm operators who began their farming careers in the wheat grain-sorghum region of central Kansas. Study analysis found land tenure arrangements used at various stages of farm development to have statistically measurable impact upon short- and long-run growth rates.

PAUL E. ROSENBERRY, Ph.D. Iowa State 1971. A proposal for improving the U.S. Department of Agriculture's cost and return studies.

This study is primarily concerned with refining and extending the methodology used by the U.S. Department of Agriculture in the cost and return by type of farm project. A computerized cost and return model was developed in the study that will compute the annual estimates of the costs and returns series and improve data generated for each of the series.

DAVID J. S. RUTLEDGE, Ph.D. Stanford 1970. The relationship between prices and hedging patterns in the *U.S.* soybean complex.

This study reviews concepts of hedging in commodity futures markets paying particular attention to assumptions about hedgers' attitudes toward risk. With this review as backdrop, hedging is then cast as a problem of decision making in uncertainty. The tools of portfolio analysis are shown to provide a means of reconciling different concepts of hedging and also to provide for a model of hedger behavior. This model is applied to the *U.S.* soybean complex.

ROBERT ST. LOUIS, Ph.D. Michigan State 1970. A simulation analysis of planned dairy farm development in eastern Quebec.

Due to serious criticisms of the use of public funds for regional development programs in milk production, the study first examined planning issues in the dairy subsector. Low resource productivity in dairying was singled out as a major source for increasing farm income. Using a simulation model, alternative strategies for increasing dairy productivity were explored. The results indicate the highest payoff strategies were large increases in high quality feed and artificial insemination.

WILLIAM W. SAITTA, JR., Ph.D. Alabama 1970. Resource allocation with special emphasis on recreation planning at water impoundments.

EDMOND E. SEAY, JR., Ph.D. Iowa State 1970. Minimizing abatement costs of water pollutants from agriculture: a parametric linear programming approach.

The study focuses on suspended sediment to typify nondegradable, diffuse-source water pollutants from the agricultural sector. Pollution abatement is viewed as a collective good. Exogenously specified water quality standards for several uses are taken as indications of societal preferences. Using alternative erosion control methods as activities in a linear programming framework, least-cost estimates of achieving the standards were obtained. Several institutional arrangements for sharing abatement costs were examined and evaluated.

ARTHUR H. SMITH, Ph.D. Arizona 1971. A socioeconomic analysis of the goals and attitudes of Arizona cattle ranchers.

An analysis of the economic and attitudinal factors affecting the decisions cattle ranchers make in light of the very low return on investment received from cattle production. The quantitative techniques of factor and discriminant analysis show that the consumption aspects of ranch ownership strongly affect the economic decision ranchers make and that the market prices of ranches are significantly higher than values in production only.

GHOLAMREZA SOLTANI-MOHAMMADI, Ph.D. California (Berkeley) 1970. Development of irrigated agriculture in Iran: the problem of choice between irrigation techniques.

A linear programming model is developed to select between modern surface and sprinkler systems in the two irrigation projects in Iran. Skilled labor is included in the model as one of the major limiting factors. Under the specified conditions and assumptions a portable or semiportable sprinkler system is indicated to be economically superior to a modern surface irrigation for the development of new lands. However, in lands where a modern surface irrigation system has been installed the transformation to sprinkler irrigation is not feasible.

BERNARD H. SONNTAG, Ph.D. Purdue 1971. Simulated near-optimal growth paths for hog-corn farms under alternative resource, price and efficiency situations.

The simulation model developed is designed to aid in planning growth, management system and building system over time. Net worth of the farm is maximized given the risk level, resource supplies, price and technical coefficients, consumption requirements and production alternatives. Output includes 1) annual farm plans; 2) a net worth statement with detail on debts and assets; and 3) biweekly flows of cash, labor, feed, hog sales and purchases, etc.

LONNIE M. SPURGEON, Ph.D. Missouri 1970. Pricing arrangements for one grade of milk.

Alternative pricing arrangements to allow for equitable distribution of proceeds from the milk market with one grade of milk were presented and analyzed. A six-class pricing system was formulated. Demand and total revenue functions were estimated. Quadratic programming was employed to find the price for each class of milk which would maximize total revenue subject to the constraints imposed. Results demonstrated possibilities of using classified pricing to maximize dairymen's revenue.

TERRY P. SUTTON, Ph.D. Kansas State 1970. Temporal stability of the effects of factors causing intertract variations in farm real estate prices in Kansas.

Interyear variations in farmland prices are reflected in changes in the influences of factors (soil quality, etc.) causing intertract differences in prices. The primary concern of this study was to identify forces causing factor influences to vary between years. It was concluded that inflation and changing technology were the significant causes of factor influence changes. Finally, an estimating equation from which to predict future prices of specific tracts was derived by multiple regression techniques.

WAYNE C. THOMAS, Ph.D. Washington State 1971. Economics of supplementary livestock production in the Columbia Basin of Washington.

Uncertain annual fluctuations in farm income are a major problem confronting the agricultural industry in the United States. A linear programming model was constructed which included crop and livestock possibilities. Using separable programming individual enterprise net income variance and covariance between enterprises were used to constrain the linear programming model. As the permissible level of net income variance was decreased, reductions in farm net income and changes in enterprise mix were noted.

H. DON TILMON, Ph.D. Purdue 1971. An economic analysis of the pork marketing channel in Indiana.

The trend by the larger chains toward central warehousing, processing and distribution of meats appears to enhance efficiency and increase competitive pressure on other retail firms. It also tends to reduce margins of packers and processors on some of their higher profit items. The demand for pork was found to be lowest in the second quarter and highest in the fourth. Demand appeared to have increased significantly between the 1955–65 and 1966–68 periods.

ALAN R. TUBBS, Ph.D. Cornell 1971. Capital investments in agricultural marketing cooperatives: implications for farm firm and cooperative finance.

The thesis used simulation as a technique to investigate the impact of alternative cooperative investment plans on the farm firm of the cooperative member. The nature and amount of investment required by the cooperative significantly limited the growth potential of the farm firm. Cooperative management must evaluate alternatives for meeting the cooperative's need for capital from the viewpoint of the effect on member firms as well as on cooperative financial structure.

GAIL E. UPDEGRAFF, Ph.D. California (Berkeley) 1970. The economics of sewage disposal in a coastal urban area: a case study of the Monterey Peninsula, California.

Although not sufficient, "benefit-cost" analysis is a necessary part of the decision process for water quality. Therefore one part of the study consists of a benefit-cost analysis of alternative systems of Peninsula sewage disposal in terms of their costs and the objective they satisfy. The second stage of the study consists of an analysis of institutional problems associated with water pollution. This included a critical review of the enforcement mechanism of and the authority granted to the state by the Porter-Cologne Water Quality Control Act.

JOHN VONDRUSKA, Ph.D. Michigan State 1971. An economic analysis of small watershed project evaluation procedures.

The project evaluation procedures used by the U.S. Soil Conservation Service were translated into a computer program and a sensitivity analysis made of the major criteria and variables. These included crop-enterprise cost and return variables and hydrological variables as well as timing, interest rate and investment criteria. Small variations in hydrologic variables, well within the range of error, had a greater effect than many of the economic parameters.

JOSEPH S. WEISS, Ph.D. Cornell 1971. The benefits of broader markets due to feeder roads and market news: northeast Brazil.

This study identifies expected costs and benefits facilitating services for agricultural markets. Improved roads and improved communication facilities were considered. Benefits include price improvement to producers and also a broadening of markets. It is concluded that broader markets will stimulate development of larger firms, ameliorate monopolistic tendencies and encourage deployment of new technology. Quantitative research techniques are used extensively in estimation of costs and benefits.

DAVID A. WHITAKER, JR., Ph.D. Florida 1971. Economic effects of trade policies on the shrimp fisheries of the United States and the Latin American nations.

This study is an empirical analysis of the effects of a free trade policy for a single product, shrimp, on factors such as income, profitability, investment, and employment in the shrimp trawling industry. A supply and demand framework is utilized, with special consideration given to the open access nature of the resource. The impact of hypothetical restrictive policies is evaluated. Major emphasis is placed on the United States fishery.

REED W. WILLIS, Ph.D. Colorado State 1970. Political economy of Colorado's ground water development and use.

Characteristics constituting a "good" ground water law were developed and comparison was made with the Colorado law.

GAYLORD E. WORDEN, Ph.D. Iowa State 1970. An interfirm competition model for deriving empirical estimates of supply response.

The primary objective was to develop a supply response model for agriculture that would functionally relate firms to the aggregate. Representative farms for Iowa were used as submodels of a state aggregate linear programming model. Resources were limited at the firm, subaggregate and aggregate levels. The interfirm competition for resources limited in the aggregate caused the supply response of firms to be conditioned by the combined actions of all firms.

DEBEBE WORKU, Ph.D. Colorado State 1971. The economics of water use: an inquiry into the economic behavior of farmers in West Pakistan.

The specific objective of this study is to examine the problem of water supply in West Pakistan. A broad survey of the existing supply and trends of water development and usage was made. The sources of the problems were identified.

KENNETH B. YOUNG, Ph.D. Montana State 1971. Optimal use of ground and surface water in the Gallatin Valley, Montana.

Intraseasonal interaction of the ground and surface water system figures prominently in this irrigation-planning study of a river basin. A model for optimizing use of variable streamflows and supplemental ground-water reserves is developed that simultaneously evaluates economic interaction occurring within the ground and surface water system during four discrete periods of the irrigation season.

JAN ZWARTENDYK, Ph.D. Pennsylvania State 1971. Economic aspects of surface subsidence resulting from underground mineral exploitation.

Manpower, Labor, and Population; including Trade Unions and Collective Bargaining

ORLEY C. ASHENFELTER, Ph.D. Princeton 1970. Racial discrimination and labor markets.

This thesis consists of three related empirical essays on the economics of discrimination against black workers. The three essays are concerned with racial discrimination and trade unionism; the interindustry structure of the wages of black workers relative to white workers; and changes in labor market discrimination over time. In each area a set of hypotheses is developed and tested using conventional econometric methods.

GAMAL BADAWI, Ph.D. Indiana 1970. Job vacancy data as an indicator of the demand for labor: an evaluation.

Exploration of the relevance of job vacancy data to labor market analysis and an evaluation of the U.S. Department of Labor's experimental job vacancy surveys.

WILLIAM F. BARNES, Ph.D. Washington (St. Louis) 1971. Wage flexibility of unemployed jobseekers.

This study investigates the transmission mechanisms through which the experience of unemployment leads to wage flexibility by the unemployed individual. The empirical observations on behavior of unemployed individuals are provided by a questionnaire survey of registrants with Employment Service offices. Regression analysis was used to test the significance of hypothesized explanatory variables of wage flexibility. Estimates of wage flexibility of the unemployed individual are used to construct a Phillips curve, supporting its micro-economic foundations.

DONALD M. BELLANTE, Ph.D. Florida State 1971. A cost-benefit analysis of the Florida Federal State Vocational Rehabilitation Program.

This is a study of the economic impact of the federal state rehabilitation program. It was conducted on cases closed in Florida during 1969. The number of clients included in the study totaled almost 14,000. Estimates were made of subgroups within the population of rehabilitants. The overall cost-benefit ratio was 31.68 and ratios for subgroups ranged from 1.80 to 86.28.

LEE K. BENHAM, Ph.D. Stanford 1970. An economic analysis of the labor market for registered nurses.

A simple econometric model is developed to investigate the simultaneous interaction of the demand for registered nurses, their labor force participation, and their geographical distribution. The parameters of this three-equation model are estimated by three-stage least squares, using cross-sectional data by states for 1950 and 1960. An investigation is also made of the determinants of the choice between registered nursing training and general college education by women.

Roger H. Bezdek, Ph.D. Illinois (Urbana) 1971. Manpower implications of alternate patterns of demand for goods and services.

Using input-output analysis this dissertation presents the development of a general, comprehensive, empirical system capable of generating detailed industrial and occupational manpower demands from alternate distributions of national expenditures reflecting different national priorities. This general system is then used to simulate the manpower effects which would be generated by various types of priority choices and expenditure patterns. A number of significant conclusions relating to the sensitivity of manpower demands are derived.

Jon Birkeli, Ph.D. South Carolina 1971. An estimate of the monetary cost of lives lost and disabled in Vietnam.

David W. Breneman, Ph.D. California (Berkeley) 1970. The Ph.D. production process: a study of departmental behavior.

Robert E. Burney, Ph.D. Florida State 1971. The role of Florida Agricultural and Mechanical University in the development of human capital.

Estimates of the amount of human capital developed by a predominantly black, state supported institution of higher learning (Florida A and M University) are derived in terms of the net differentials in lifetime income generated. The study covers graduates of the university between 1958 and 1969 and presents data on per student costs of education (public and private) as well as information on earnings. Study includes breakdowns by school and college within the university.

Gerald F. Cavanagh, D.B.A. Michigan State 1970. Black and white workers' attitudes in three industrial plants: view from the grassroots.

An interview study of 300 workers, managers, and community leaders revealed that black workers are slightly less satisfied with their jobs, boss, promotion opportunities, and relations with fellow workers. Blacks do good quality and quantity of work. Younger black males have poorer absenteeism and tardiness. Although most workers got along well with foremen, these men did not as often have good relations. Foremen of disadvantaged workers must have: personal interest, two-way communication, honest discipline.

Corindo J. Cipriani, Ph.D. Minnesota 1970. A labor market model: married women.

This study tests a simultaneous-equations model, including separate supply and demand schedules, representing the labor market for married women. Key findings are 1) that wage-elasticities of demand are about the same size as those of supply, suggesting that the widely used single-equation supply models are incorrectly specified; and 2) that a satisfactory explanation of the secular rise in the labor force participation rate of married women has yet to be developed.

Edward M. Cross, D.B.A. George Washington 1970. The behavioral styles, work preferences and values of an occupational group: computer programmers.

A job analysis of the computer programming occupation was conducted, using the constructs of a self-report personality inventory, the Job Analysis and Interest Measurement (*JAIM*), as a basis for infering desirable worker traits. The *JAIM* was then administered to computer programmers to confirm the job analysis. It was concluded that *JAIM* constructs have substantial value for conducting job analysis.

William A. Dawson, Ph.D. Wisconsin (Madison) 1971. Trade union developments in western Indian textiles.

The dissertation traces the development of the Textile Labor Association of Ahmedabad and the Rashtriya Mill Mazdur Sangh of Bombay from raw protest movements to formal complex organizations, soundly financed from internal resources and competently and adequately staffed. It is proposed that they have followed a "normal" course of evolution in response to the needs and logic of the industrial situation, although their strong institutionalization contrasts strongly with the prevalent pattern of trade unionism in the country.

Bruce F. De Vine, Ph.D. Claremont School 1970. The *U.S.* student exchange program: reverse foreign aid?

The United States, despite its provision of higher education resources for vast numbers of students from developing countries, may gain financially on student exchange with these countries, primarily because of student nonreturn. Cost-benefit analysis is applied to data for the academic year 1964–65. The calculation of costs indicates a net *U.S.* cost significantly lower than that usually associated with the exchange program. The hypothesis is upheld at very low discount rates for human capital.

Daniel T. Dick, Ph.D. Claremont School 1971. Race as a factor of production in the *U.S.* manufacturing sector, 1940–1960: an econometric approach.

A production model for manufacturing is estimated based on the hypothesis that production relationships are of the Mukerji type. A two-stage estimation procedure, necessitated by the simultaneous nature of the system, is performed on two-digit SIC manufacturing data. Results show that marginal productivity of blacks is less than that of whites holding education constant; that derived demand for black labor is more wage elastic than that for white labor; and that blacks substitute across educational levels with much more ease than do whites as shown by the constant ratios of elasticities of substitution.

Gerald F. Downey, Ph.D. Boston College 1971. A cost-benefit study of vocational education in Haverhill, Massachusetts.

This study compares the benefits and costs of vocational and regular high school education in Haverhill, Massachusetts. Substantial differences in average per pupil cost and benefits were found to exist between the two types of training. The conclusions of the study suggest that under liberal investment assumptions the vocational program as a whole is a worthwhile public investment.

Barry L. Duman, Ph.D. Southern California 1970. Participation of rank-and-file member in the decision-making process and his access to the grievance machinery of his union: including a survey and case study of UAW Local 1531, Amarillo, Texas.

This study analyzed two areas of current importance. The first pertained to the character of the American labor union with respect to democracy as a vital ingredient. The second dealt with the attitudes of the rank-and-file worker as he relates to the various internal aspects of his union. Additionally, since it was found that members are universally concerned with the provisions of the grievance procedure, recent legal developments in this area were analyzed.

Zachary Y. Dyckman, Ph.D. Pennsylvania 1971. An analysis of Negro employment in the building trades.

Fifteen cities were visited during 1968 to determine the extent of racial discrimination in construction. After much pressure, Negroes were gradually being admitted into unions that until recently excluded them entirely. Theoretical models were developed to explain the consequences of discrimination among both unionized and nonunionized construction workers. To test the implications of the theory and other hypotheses, 1960 census data on employment, earnings and unemployment of white and Negro workers were used.

Linda N. Edwards, Ph.D. Columbia 1971. Investment in human capital: a study of the teenage demand for schooling in the United States.

An economic model of the determinants of the school enrollment of teenagers (fourteen to seventeen year olds) is developed and estimated with two bodies of cross-sectional data for 1960: data for 426 congressional districts and data for 48 states. The cross-sectional results are then used to explain the secular growth in teenage enrollment rates between 1920 and 1960. Cyclical variations in teenage enrollment rates for the 1948 to 1967 period are also analyzed.

Ronald G. Ehrenberg, Ph.D. Northwestern 1970. The short-run employment decision and overtime behavior in *U.S.* industry, 1966.

A theoretical and empirical study of the employment and hours decisions of firms. The empirical results are used to derive estimates of the impact of changing the overtime premium wage on employment and overtime hours in *U.S.* industry.

Edward C. Erickson, Ph.D. Southern California 1970. An economic analysis of the factor market for pharmacists.

This study applies economic terminology and theoretical analysis to the factor market for pharmacists in an effort to identify the supply and demand conditions and influences on the factor market according to their impact on supply and demand. An evaluation of the factor market's performance is made in qualitative terms by citing evidence on dynamic patterns within the context of the market's structure.

Charles K. Fairchild, Ph.D. Duke 1971. Subsidized worker relocation in the United States.

Ghazi M. Farooq, Ph.D. Pennsylvania 1970. Dinensions and structure of labor force and their changes in the process of economic development: a case study of Pakistan.

Since no comprehensive work in the labor force field has been done for Pakistan, this study is equally divided between the dimensional and the structural aspects of labor force. It has both substantive and methodological focuses and, with appropriate modifications, can be used as a prototype for labor force analysis in developing countries in general.

Terry G. Foran, Ph.D. Pennsylvania State 1971. The effect of market structure on the interindustry wage differential in *U.S.* manufacturing, 1947–58.

Kenneth L. Gagala, Ph.D. Michigan State 1970. Racism in the building trades.

The problems of low black participation in the building trade unions was investigated by interviewing civil rights government, industry and union representatives in Chicago and Detroit and analyzing data provided by them. The major obstacle is the paucity of blacks with the requisite skills for working in the trade. Remedial action depends upon the power the black community exerts on government and the unions.

Kenneth E. Galchus, Ph.D. Washington (St. Louis) 1970. The elasticity of substitution of white for nonwhite labor.

It is shown, using both white and nonwhite Census data for twenty-one occupational categories, that whites and nonwhites are perfect substitutes in production once variations in the relative levels of educational attainment have been eliminated. This evidence indicates that there is very little Becker-type discrimination. Moreover, the existence of straight line isoquants indicates that much of the discriminatory differential, which Becker attributes to racial discrimination, can actually be attributed to quality differentials.

Lynn C. Griffiths, Ph.D. Colorado 1970. Some implications of an investigation of the part-time labor force in the United States to the theory of hidden unemployment.

John W. Hambleton, Ph.D. Wisconsin (Madison) 1971. Determinants of geographic differences in the supply of physician services.

The purpose was to discover how effective in explaining the per capita distribution of specialists, general practitioners, and graduate trainees by state, county, and postal zone are characteristics reflecting current wages and the costs of using leisure time. Linear, ordinary least squares regression indicated that prospects for leisure influence specialists more than general practitioners, wages have the opposite effect, and neither influence graduate trainees noticeably.

JEFFREY L. HARRISON, Ph.D. Florida 1970. Union discipline and its implications for the individual and for management.

In an effort to determine how much freedom unions have in disciplining their members, labor legislation, union constitutions, NLRB and *U.S.* court decisions and decisions by arbitrators were studied. Unions have had and continue to have wide disciplinary powers which may affect the employer-employee relationship in a very direct sense. In addition, union disciplinary actions are frequently void of many procedural safeguards.

MASANORI HASHIMOTO, Ph.D. Columbia 1971. Factors affecting state differences in unemployment.

The dissertation analyzes differences among states in cyclical fluctuations, levels, trends, and seasonal fluctuations of unemployment. The time-series unemployment data (1950I–1968IV) are decomposed by regression techniques into these components, which are analyzed in terms of variables suggested by the economics of human resources. The findings are related to issues of poverty, income inequality, and depressed areas. Work experience is a major factor related to these problems. Yet the cross section unemployment rate is found to be a poor measure of work experience, especially for nonwhites.

JAMES J. HECKMAN, Ph.D. Princeton 1971. Three essays on the supply of labor and the demand for goods.

The first essay examines the effect of exogenous and endogenous wage growth on household life cycle labor supply and consumption decisions. Contrasts between the empirical implications of static and dynamic models are developed. The second essay proposes and implements new methods of estimating household demand parameters. Estimates of household labor supply equations generally consistent with a priori economic theory are presented. The third essay considers the impact of effective time restrictions on household economic behavior.

JAMES A. HEFNER, Ph.D. Colorado 1971. Black employment in a southern "progressive" city: the Atlanta experience.

DAVID C. HERSHFIELD, Ph.D. Princeton 1970. Automation and collective bargaining in the New York City newspaper industry.

This dissertation examines policies of three unions in terms of Slichter's theories. Conclusions: typographers chose between policies of control and obstruction according to effects on skills and labor-displacing power. Photoengravers and stereotypers were too weak to practice other than policy control. Five daily newspapers failed primarily from lack of advertising and circulation, not rising labor costs. Typographers' contract demands emphasized "automation-related" benefits, as measured by an original formula. "Share of savings" approach has merit.

C. RUSSELL HILL, Ph.D. Minnesota 1970. The economic determinants of labor supply for the urban poor.

This study determined the effect of a variety of economic and demographic variables on the quantity of labor supplied by white and black male heads of poor families. The findings suggest that there are statistically significant differences in the effect of wage rates, education, health, and several other variables on the quantity of labor supplied by white and black poor family heads and on the quantity of labor supplied by poor and nonpoor heads.

ROBERT J. HINES, Ph.D. Cornell 1970. An econometric model of external labor supply to the establishment within a confined geographical market.

CHRISTINE L. HOBART, D.B.A. Harvard 1971. A study of the major influences on nurses' collective bargaining at selected Massachusetts hospitals.

This descriptive analysis of nurses' collective bargaining over a four-year period indicated that in spite of a comparatively weak form of collective bargaining, the registered nurses made substantial economic gains in the period. The causes for this weakness lay for the most part in the history of the development of the nursing profession. Whatever power the nurses had was derived for the most part from the unique economic system of the health care industry. In the light of this analysis two modes for future nurses' bargaining were projected and evaluated: a professional model and a guild-union model.

STANLEY W. HUFF, Ph.D. Princeton 1971. Hiring practices, hiring standards, and wage differentials for laborers and clerical workers in a small labor market.

This study analyzes the recruitment, selection, and compensation of laborers and clerical workers in a relatively small labor market. It uses parametric and nonparametric tests to assess the impact of wage and nonwage forces on worker quality and to explore the interdependencies which exist. The results suggest that economic factors partially explain behavior in the market for these workers but also show the need for further research in this area.

WILLIAM G. JOHNSON, Ph.D. Rutgers 1971. The influence of health on labor force participation.

The conventional labor force participation model is expanded to include proxies for "health." Data from a national sample of chronically ill persons are used to test the model. The causes of observed differences in black-white participation are clarified. "Age discrimination" in hiring is estimated. The work-behavior of family members who are not ill and its effect upon the sample respondent is analyzed and estimated, as is the impact of public and private transfer payments.

ABRAHIM KAMEL, Ph.D. Michigan 1970. The impact of Nasser's regime on labor relations in Egypt.

William J. Kelly, Ph.D. Rice 1971. A cost-effectiveness study of clinical methods of birth control.

The study develops a method for calculating the marginal cost of preventing a birth with different method-user combinations. Innovations include production functions for contraceptive services, continuation functions which include age, Poisson patient arrival functions, a method for estimating the number of births prevented using parity data, and procedures for determining optimum clinic capacity under queuing conditions. Empirical results for Puerto Rico cover five different methods administered to patients of ages 20 to 40. Finally, recommendations are made for operational improvements and further research.

Duane A. Kujawa, Ph.D. Michigan 1970. International management of labor relations in the major American automotive corporations: a comparative study of decision-making patterns and their determinants.

Ronald D. Lee, Ph.D. Harvard 1971. Econometric studies in topics in demographic history.

A model for population projection that incorporates some of the economic consequences of demographic variables is developed and applied in this work. Economic and demographic data from preindustrial England are used to test and estimate forms of simple analytic models, that are couched in basic theoretical issues. The basic model is then extended to the economic growth of the United States and its application to this country's population in the twentieth century.

Hartley Lewis, Ph.D. Rochester 1971. The importance of turnover costs in the male-female wage differentials.

Using individual data from the 1960 U.S. Census, male-female wage differentials are obtained for two digit manufacturing industries. These differentials are combined with differences by sex in turnover rates to obtain estimates of the specific investment of employers in employees. While this investment is substantial, the difference by sex in the turnover costs arising from its loss explains only a small fraction of the wage differential.

Charles R. Link, Ph.D. Wisconsin (Madison) 1971 The quantity and quality of education and their effects on the earnings of chemical engineers.

The study analyzes the determinants of chemical engineers' earnings with emphasis placed on the association between earnings and the quality and quantity of education. The empirical inquiry is based largely on cross-sectional data from the Postcensal Survey of Scientists and Engineers. The relationship between earnings and several educational, socioeconomic, and demographic characteristics is measured simultaneously through coefficients of a multiple regression operation.

Luis R. Luis, Ph.D. Notre Dame 1970. An economic analysis of metropolitan migration flows in the United States.

The determinants of aggregate migration flows to and from each of fifty SMSAs are studied by means of a two-equation model. The main conclusion is that in and out gross migration flows for SMSAs are simultaneously determined. The relationship between migration and the SMSA's productive structure is also analyzed. Aggregate production functions for each of the fifty SMSAs are estimated to measure migration's contribution to SMSA economic growth.

D. Bruce McKinlay, Ph.D. Oregon 1971. A functional classification of occupations.

William F. X. Maher, Ph.D. Boston College 1970. A case study of the shortages of general duty nurses in metropolitan Boston hospitals, 1956–1964.

Michale J. Mestrovich, Ph.D. Notre Dame 1970. The contributions of MDTA to an urban economy: a case study.

The study analyzes the impact of the Manpower Development and Training Program primarily as a means of investing in human capital and secondly as a method of alleviating structural mismatches in urban labor markets. Multiple regression analysis utilizing cross-sectional time-series data on the socioeconomic characteristics of trainees participating in the program showed that MDT significantly affects the wage rate and job opportunities of graduates. The program was also shown to lessen structural inequalities in the labor market.

James R. Miller, D.B.A. George Washington 1970. Selecting computer programmers: a multivariate approach to the determination of predictors using an improved criterion for on-the-job success of male and female computer programmers.

Research had three objectives: determine population difference between male and female computer programmers in relation to selected variables; determine if selected variables are statistically valid for predicting on-the-job success; and develop a subjective rating instrument for evaluating computer programmer performance. Results indicated no population differences at the one percent level of significance. Abstract reasoning and quantitative reasoning aptitude are moderately predictive. A rating instrument was developed for use of programmer supervisors.

Thomas J. Moore, Ph.D. North Carolina 1970. Workmen's compensation in North Carolina.

The dissertation represents a study of the workmen's compensation laws of North Carolina as contrasted with other states. The subject of workmen's compensation has been neglected and little has been done in focusing attention to the inadequacy of the laws dealing with occupational diseases and monetary benefits. The procedures and administration of the act are complex and little has been done to acquaint the worker of his rights and benefits under the law.

James R. Nininger, Ph.D. Michigan 1970. College graduate turnover in industry: a study of the role of expectations in decisions to terminate or continue participation in an organization.

JOHN F. O'CONNELL, Ph.D. Wisconsin (Madison) 1971. The labor market for engineers.

A two-equation model of the engineering labor market is developed and estimated via ordinary least squares and two-stage least squares. Both cross section data by state and time-series data by industry for the period 1953 to 1968 is used. The model is applied to various engineering specialties and a comparative analysis of the determinants is made. Particular attention is given to estimating the wage elasticities of demand and supply.

CHARLES A. ODEWAHN, Ph.D. Kentucky 1970. The contract rejection in Kentucky.

An investigation of four major hypotheses proffered to explain the phenomenon of contract rejection by union memberships. Management and union negotiators and federal mediators were questioned about the importance of social, legislative, structural, and economic factors in specific negotiations which resulted in membership rejection. A theoretical model is developed which supports a relationship between rejection and general business conditions. The conclusions drawn from survey data support this explanation.

ELIOT S. ORTON, Ph.D. Cornell 1971. Inexperienced workers in the labor market.

DONALD O. PARSONS, Ph.D. Chicago 1970. Specific human capital: layoffs and quits.

This study formally derives a number of propositions about the relationship of firm specific human capital and labor turnover. The dynamic implications are examined empirically using time-series data for layoff and quit rates in selected manufacturing industries. The static hypotheses that layoff rates and quit rates are negatively related to the firm's and worker's shares respectively of specific capital is tested using cross section manufacturing data. The implications receive strong empirical support.

RICHARD D. PAYNE, Ph.D. Southern California 1970. An economic analysis of nurse mobility patterns.

A survey design to distinguish between the shortage and the mobility of registered nurses. A random sample of twenty hospitals in Los Angeles tabulated all nurses being hired and terminated in one year. Almost 1500 nurses were interviewed, information was coded and tabulated for computer processing. Wage-induced geographic mobility and turnover rates were low. Few changes were job related or wage related; most were catalogued as "beyond the nurse's control." A monopsony hypothesis lacked support.

LEWIS PERL, Ph.D. California (Berkeley) 1970. The role of educational investment and family background in the production and distribution of educational services.

JERRY E. POHLMAN, Ph.D. Cornell 1970. Job creation and welfare: a trade off analysis.

PETER A. PROSSER, Ph.D. Cornell 1970. Concentration and the rate of change of wages in the United States, 1950–1962.

RAYMOND V. RASMUSSEN, Ph.D. California (Berkeley) 1970. Exchange and equity in the task team: testing a model of social interaction.

STANLEY R. REBER, Ph.D. Chicago 1970. The effects of unionism upon relative wages.

The study attempted to estimate empirically by regression techniques the effects of unionism upon relative wages in each of some seventy industries during the period 1945 to 1968 in the United States. Data contained in the *BLS* Industry Wage Surveys of this period were analyzed. Estimates of the global effect of unionism upon *U.S.* relative wages for each of the years 1945 to 1968 were built up from the underlying industry estimates.

IRA J. REES, Ph.D. Alabama 1970. The role of the attorney-consultant in labor relations in a southern metropolitan area.

SUN WHAN RHEE, Ph.D. Duke 1971. Welfare economics and the theory of optimum population: a reformulation of welfare notion in optimum population.

ROBERT L. RITTENOURE, Ph.D. Texas (Austin) 1970. Negro employment in the federal service in the South.

Negroes are relatively underrepresented in federal service white-collar and relatively overrepresented in blue-collar occupations in the South. The upgrading of blacks in both white- and blue-collar occupations has not been comparable to that for whites. In the postal service blacks have done much better, both with respect to penetration and upgrading, than in other areas of the federal service. Programs designed specifically for the South, such as independent evaluation and complaint adjudication, are necessary if the federal government is to become an equal opportunity employer in the South.

MYRON ROOMKIN, Ph.D. Wisconsin (Madison) 1971. An evaluation of adult basic education under the Manpower Development and Training Act in Milwaukee, Wisconsin.

This study evaluates the adult basic education program, i.e., elementary educational instruction for adults financed under the MDT Act in Milwaukee, Wisconsin. The study focuses on the increased average annual earnings benefits of the program, but other noneconomic benefits are examined as well. Employing two alternative research designs, a similar evaluation of basic education is produced: benefit-cost ratios show basic education as an inefficient allocation of public revenue.

BERNARD D. ROSTKER, Ph.D. Syracuse 1970. Manpower theory and policy and the residual occupational elasticity of substitution.

RICHARD M. SCHEFFLER, Ph.D. New York 1971. An empirical investigation into the geographic distribution of physicians and specialists.

This study describes and analyzes the geographic distribution of physicians and specialists in the United States from 1956 to 1967. A macro-economic location model is theoretically developed and estimated. The model includes economic, demographic, and institutional variables such as hospital facilities, medical schools, population, per capita income, and state licensing figures. Policy recommendations are suggested to improve the spatial distribution of health manpower in the United States.

CHARLES E. SEAGRAVE, Ph.D. Stanford 1971. The southern Negro agricultural worker: 1850–1970.

This study estimates the changes in real incomes to Negro field hands resulting from their emancipation. Wage and income estimates are developed for freedmen in the state of Louisiana from original payroll records. The incomes are compared with the value of the estimated maintenance goods received by male field hands prior to the War. The comparison shows significant income gains to prime field hands resulting from their emancipation.

DAVID O. SEWELL, Ph.D. Duke 1971. Training the poor: a benefit-cost analysis of vocational instruction in the U.S. Antipoverty Program.

ALLEN W. SMITH, Ph.D. Indiana 1970. Collective negotiations in the public schools: Indiana, a case study.

This thesis determines the present status and the economic and social consequences of collective negotiations in the public schools of Indiana. A comprehensive questionnaire was sent to each of the 287 school corporations in the state. In addition, an in-depth study was made of a small group of school systems, selected on the basis of the data received from the total population study.

JOHN L. SNARE, Ph.D. Duke 1971. An economic analysis of the veteran's education program below the college level: 1945–1955.

LEE A. SODERSTROM, Ph.D. California (Berkeley) 1971. The demand for labor in residential construction.

WILLIAM E. SPELLMAN, Ph.D. Kansas State 1970. Projections of occupational requirements for Kansas in 1980.

The objective of the study was to quantify occupational structure of the state and to develop a methodology for forecasting occupational requirements. An input-output model of the economy was utilized to forecast output levels for each of the fifty-five industrial sectors. Employment levels were projected on basis of productivity-adjusted levels of output. A best-practice model was used to project occupational mix. This allowed for inclusion of both technological change and interindustry production structure.

MELVIN L. STONE, Ph.D. Colorado 1971. A spectral-analytic test of the wage-lag hypothesis.

DONALD H. STRASZHEIM, Ph.D. Purdue 1971. The income performance and occupational structure of college graduates.

The literature on the economics of higher education deals with many questions. Lack of reliable data, however, has left questions distinguishing between character of academic training largely untouched. This dissertation makes use of a new data source in which bachelors degree holders are subdivided by academic curriculum, determines income profiles and occupational distributions by academic curriculum, and attempts to measure these differences and to explain the patterns found by the nature of the academic training.

ROY D. THOMAS, Ph.D. Cornell 1970. The adjustment of displaced workers in a labor-surplus economy: a case study of Trinidad and Tobago.

JOHN M. WALKER, Ph.D. California (Berkeley) 1970. Models of academic governance in emergent universities.

JOHN P. WALTER, Ph.D. Notre Dame 1971. The economics of labor force participation of urban slum-barrio youth in Cali, Columbia: a case study.

The dissertation studies the importance of youth in the economic development process. While it considers economic factors which influence the labor-force status of youth, its main purpose was to consider noneconomic factors which influence youth in their acquisitive behavior of enrolling in school or entering the labor force. These factors were attitudes and the psychological influences of the youth's parents. Contrary to major landmark works by Everett Hagen and David McClelland, the study concludes that the father plays a very active role in influencing the youth's acquisitive behavior in a rejected urban slum society.

HOWARD J. WILLIAMS, Ph.D. Pennsylvania State 1971. The treatment of population as an economic factor.

DONALD WISE, Ph.D. Claremont School 1971. Bracero labor and the California farm economy: a micro study of three crops, 1952 through 1967.

An eight-equation econometric model was used to estimate the impact of foreign contract labor on the harvest labor market in California. The wage-elasticity of supply for domestic workers was found to be relatively high. This result contradicts the "stoop labor" hypothesis used by growers to justify the importation of foreign workers. In addition, the use of foreign workers was found to have little effect on the produce market.

MELVILLE Z. WOLFSON, Ph.D. Illinois (Urbana) 1971. A reexamination of the wage and employment effects of the minimum wage on the southern pine industry.

Welfare Programs; Consumer Economics

ALEXANDER V. ALEX, Ph.D. Indiana 1971. Costs and returns on investment in education: a case study, India, 1950–61.

The issue raised is whether education played a productive role in Indian economy during the 1950's. Resource cost valuation of education investment in

India shows that it increased at a higher rate than national income during the period 1950–61. The stock of education increased in relation to physical capital. This raised the economic quality of the Indian labor force. Benefits of education were estimated from life earnings of educated persons and returns to education were estimated.

David M. Ambrose, D.B.A. George Washington 1971. A study of marketing fish protein concentrate enriched products for preschool children in Latin America.

This study examined the importance of seventeen environmental factors to the marketing of fish protein concentrate products to protein-malnourished preschool children. The Delphi method using questionnaires to forty-six experts established the importances. The participation of the institutions of the environment, parental attitude toward childhood nutrition, and family income were the most important factors. Attitude toward fish, availability of local food stores, and the literacy level of the adults were among the least important.

Robert G. Beck, Ph.D. Alberta 1971. An analysis of the demand for physicians' services in Saskatchewan.

Kenneth R. Biederman, Ph.D. Purdue 1971. Tax credits and their effects upon the employment and training of the economically disadvantaged.

This study presents a theoretical-empirical analysis of the effects of generalized, hypothetical employment-training tax credit programs upon the demand for the unskilled, unemployed poor. The results indicate that such tax credit programs can be effective devices toward accomplishing short-run goals of hiring and training economically disadvantaged and unemployed members of the labor force. Program cost estimates are presented as well as estimated credit program effects upon related welfare programs.

Kelly J. Black, Ph.D. Southern California 1970. The size distribution of income and its relationship to economic and social well-being in the state of California.

The hypothesis of an inverse relationship between income inequality and aggregate well-being was tested using data for the fifty-eight counties of California. Gini's inequality coefficient was calculated from combined census and income tax data. Well-being was measured through factor analysis, including economic, social, and time dimensions. Little overall correlation was obtained ($r=-0.312$). Agricultural counties, however, scored lowest on the well-being scale and very high in income inequality.

Lal C. Chugh, Ph.D. Harvard 1971. An input-output model of education in the United States.

The dissertation examines critically the existing approaches to the economics of education and develops a new model, using input-output analysis, for an expanding and a stationary economy. The model is applied to three educational categories—elementary, secondary, and higher—in the United States and the final demand for the model has been generated from the foreign trade sector of the economy for 1960.

Charles W. Clifton, Ph.D. Michigan 1970. Toward a model of the housing market.

Many different problems have prevented the development of a comprehensive model of the housing market. Some of these are practical problems encountered in the analysis of housing market data. Others stem from the absence of an adequate theoretical model of the housing market. This paper addresses problems in each of these areas. In attempting to solve the practical problems, a heavy reliance is placed upon empirical analysis of data obtained in interviews conducted by the Survey Research Center with 3,039 families.

William W. Cyzewski, Ph.D. Alabama 1970. Comparative study of the financing patterns of the white and Negro low income groups in Tampa, Florida.

Neville J. Doherty, Ph.D. Michigan State 1970. Performance of the medical industry in the Grand Traverse region of Michigan.

This dissertation describes the financial structure of the Traverse City medical industry as it serves the primarily rural residents of a seven-county area. Sources and dispositions of revenues were determined. Financial drains due to inefficiencies and monopolies were estimated.

Elizabeth F. Durbin, Ph.D. Columbia 1971. Family instability, labor supply and the incidence of aid to families with dependent children.

Equations were developed and tested relating to family instability and the demand for the public assistance programs for mothers and children deprived of a male supporter and subsequent effects on labor supply. The results were successfully applied to predict public assistance incidence and male and female labor force participation rates in New York City. Policy implications for universal income maintenance programs and their effects on work incentives and family formation are discussed in detail.

Robert B. Fischer, Ph.D. Oregon 1970. Individual demand for Ph.D. degrees in American colleges and universities, 1958–1968.

Bernard Friedman, Ph.D. Massachusetts Institute of Technology 1971. Economic studies of health insurance in the United States.

Haynes C. Goddard, Ph.D. Indiana 1970. A study in the theory and measurement of benefits and costs in the public library (a theoretical and econometric analysis with special reference to Indiana public libraries).

The thesis undertakes an economic analysis of the relevant supply factors and functional relationships which determine and explain the circulation of books and associated costs for the public library system of Indiana for the year 1967. Also, an exploratory effort is made to establish a methodology which could be used to perform a benefit-cost analysis of the library circulation.

Delmar G. Good, Ph.D. Illinois (Urbana) 1970. Benefit-cost analysis of Illinois' high school driver education program.

This analysis compares the measurable costs of Illinois' high school driver education program to the stream of measurable benefits derived from the program. An effectiveness model was based on the 1963 study published by the Secretary of State of Illinois and reinterpreted that data to estimate accidents prevented in 1962. Benefits were estimated, and redistributional effects of the program identified. Significant unmeasured benefits are required to clearly justify the program.

Harry J. Halley, D.B.A. George Washington 1971. An analysis and evaluation of the *OEO* national antipoverty planning process.

Using the case study method and contemporary planning process theories, the *OEO* national antipoverty planning process was analyzed and evaluated to determine the extent to which its design was congruent with contemporary theories and the lessons that might be learned. The study revealed that the design had been congruent in varying degrees and suggested that future designers should plan such a process comprehensively, review the literature, and use an idealized model with real world specifications.

John J. Hardesty, Ph.D. California (San Diego) 1970. An empirical study on the relationship between poverty and economic prosperity.

Poverty, as officially defined, has been declining over the postwar period. The purpose of this study is to determine the degree to which economic prosperity has been responsible for the reduction, to make forecasts of the future magnitude of poverty based on this historical relationship, to ascertain the sensitivity of these findings to the particular definition of poverty utilized, and to consider the implications of the findings, both with regard to economic policy and to various aspects of economics methodology.

James G. Harris, Ph.D. Oregon 1970. Curricula costs in Oregon community colleges.

Louis H. Henry, Ph.D. Notre Dame 1970. The impact of Medicare and Medicaid on the supply and demand conditions of nursing homes.

The supply analysis deals with the number of beds per state in 1967. Poor quality conditions were found to be so important that supply data do not reveal the true level of service. A multiple regression technique was used to test the determinants of demand. Overall, the programs have been beneficial to the industry while the plight of the patient has improved little. The needs of long-term care are ignored.

Daniel B. Hill, Ph.D. Purdue 1971. An economic analysis of health insurance, with special reference to Blue Cross.

The study examines the economic impact of health insurance through both theoretical and empirical analyses. The conceptual work develops characteristics of an optimal health insurance plan. Empirically, the applicability of the optimal plan concept in evaluating a Blue Cross Plan is explored and the assumptions underlying the optimal plan are tested. Implications for national health insurance proposals are discussed.

William H. Holley, Ph.D. Alabama 1970. Employer practices relating to the hard-core jobless: a study of the metropolitan Birmingham area.

Marjorie H. Honig, Ph.D. Columbia 1971. The impact of the welfare system on labor supply and family stability: a study of female heads of families.

The study finds a significant relationship between the proportion of the population receiving public assistance payments under the Aid to Families with Dependent Children Program and the level of income available to recipients from the program relative to expected earnings. The response to increases in welfare income assumes two forms: increases in the proportion of the eligible female head of family population applying for AFDC assistance, and increases in the eligible population itself due to welfare-induced desertions.

Lien-fu Huang, Ph.D. Rochester 1971. Measurement of the effect of health insurance on the demand for medical care.

The object of this paper is to devise means to detect and measure the effect of health insurance on the demand for medical care. Under the assumption that insurance premiums accurately reflect actuarial costs, a theoretical model of consumer choice was made; the technique of probit regression analysis was applied; a sample of 11,063 households drawn by *BLS* in 1960 was examined; and some welfare policy implications were discussed.

Elise G. Jancura, Ph.D. Case Western Reserve 1971. The economic implications of the use of labor resources by hospitals.

An assessment of labor's effect on hospital costs with analysis of factors influencing the use and subsequent cost of labor resources including: changes in compensation levels, changes in the professional and departmental structure of the labor force, changes in the size of the staff relative to changes in output, and changes in the ratio of labor to nonlabor expenditures. Also studied is the variation in labor usage and cost as hospital size varies.

Ahmed Khurshid, Ph.D. Pennsylvania 1970. A comparative analysis of the British and American social security systems.

Kenneth S. Lyon, Ph.D. Chicago 1971. Some of the economic effects of subsidizing rural electrification.

This study estimates the cost to the economy of using the distorted input combinations that result from subsidies to rural electrification. These costs are the difference between the estimated social costs of the subsidized cooperatives and the estimated social costs of the investor-owned utilities for producing the quantity of electricity sold by the cooperatives during a particular time period. Cost functions, generated from cost minimization subject to a production function constraint, are used to estimate the costs. The cost to the economy for the year 1956 (a typical year) is estimated to be approximately equal to the loan funds advanced to the cooperatives by the Rural Electrification Administration during the year.

T. DAVID MCCULLOUGH, Ph.D. Cornell 1970. Unit pricing in supermarkets: alternatives, costs and consumer reaction.

Unit pricing is an example of the many policy proposals designed to protect consumers. This study identifies levels of costs and the consequences of unit pricing on consumer purchase patterns. In addition, the extent of consumer perception of the experimental system was observed along with their general attitudes toward it. It doesn't cost much, nor does it affect consumer purchase patterns. Consumers like it, however.

FREDERIC C. MENZ, Ph.D. Virginia 1970. An economic analysis of disease control programs.

EDYTHE S. MILLER, Ph.D. Colorado 1970. The concept and measurement of poverty in the United States.

JAN M. NEWTON, Ph.D. Oregon 1971. An exploratory study of the purchasing behavior of low income families.

JOHN T. O'CONNOR, Ph.D. Notre Dame 1971. Medicaid: a study of its effects on income distribution in 1968.

This study estimates the redistribution of income which was caused by the Medicaid program in calendar year 1968. The taxes used to finance Medicaid and two aspects of expenditure incidence (i.e., the direct benefits to recipients and an estimate of the inflation Title XIX caused within the medical sector of the economy) are allocated to various family income classes. The study concludes with policy considerations derived from an analysis of these patterns of income redistribution.

DONALD W. PEARSON, Ph.D. Texas (Austin) 1970. The Chilean social security system: a study of its structure and its role in the Chilean economic environment.

The dissertation is an analysis of the structure of the Chilean social security system. It examines several aspects of its operation, summarizes some of the consequences for economic development of the methods of financing benefits, and investigates social security's effects on labor force participation and consumption patterns. It summarizes recent efforts to reform the system and speculates as to the possibilites for success of future reform efforts.

DEAN O. POPP, Ph.D. Purdue 1971. Capital formation by expenditures on formal education, 1950 and 1960.

This study measures the flow of new human capital formed by elementary and secondary education in the United States in 1950 and 1960. The capital formation is disaggregated for urban and rural areas within each state. A model is then proposed to analyze differences in capital formation among these areas. Various social, economic, and demographic characteristics are used to explain different levels of capital formation.

MORGAN O. REYNOLDS, Ph.D. Wisconsin (Madison) 1971. Crimes for profit: the economics of theft.

The theoretical analysis of criminal activity identifies key variables within a "common property resource" framework. A recursive model is applied to a 1960 cross section of states and, generally, coefficients for probability of conviction, wealth, and opportunity cost prove most potent in determining the level of property crime. Punishment, regardless of how measured, has consistently negative but small effects on crime. From a policy viewpoint, however, cost calculations support increasing sentences rather than raising the probability of conviction to achieve marginal reductions in the level of crime.

ROBERT J. STAAF, Ph.D. Temple 1971. Student choice: a theoretical analysis of higher education with some empirical evidence considered.

This dissertation is an application of the axiomatic choice model from economic theory to student curriculum choice. The consumer is the student and the field of choice is comprised of alternative "course" bundles. The attainable set is defined by a student's relative aptitudes for different curricula. A "limited student sovereignty model" is developed initially. Income, grades, and course requirements are introduced as constraints. The empirical analysis suggests a substitution (high relative aptitude) and income (high or low total aptitude) effect similar to traditional demand analysis.

CAROL F. VAN ALSTYNE, Ph.D. Duke 1971. The economics and politics of low income housing.

RICHARD B. WESTIN, Ph.D. Minnesota 1971. An echo theory of automobile demand.

Consumers usually replace specific durable goods only at lengthy time intervals. This thesis examines the implications of this spasmodic adjustment process for empirical studies of the aggregate demand for durable goods over time. The model uses a two-step aggregation procedure where the number of new durables purchased and the average expenditure per new durable are estimated separately. Empirical results are presented for the demand for the new automobiles.

Urban and Regional Economics

EVA L. ATKINSON, Ph.D. Kansas 1970. Kansas City's livestock trade and meat packing industry, 1870–1914: a study in regional growth.

VIRGINIA B. BENSON, Ph.D. George Washington 1971. Growth and scale effects on employment in medium-size cities.

This study hypothesized four effects—population and per capita income growth and population and per capita income scale—on employment growth in the same urban labor market area. Multiple regression analysis is then used to test for the strength of these four effects on employment growth in six industrial groupings, over a sample of seventy-nine medium-size cities. The effect of population growth is found to be the strongest, with per capita income growth and population scale also providing some additional stimulus to local employment growth.

NELSON L. BILLS, Ph.D. Washington State 1971. A comparative study of public investments in three rural Washington communities.

This study described and analyzed municipal and school district investments in three communities that were growing, declining, and stable in terms of population between 1930 and 1965. The growth community made relatively more use of bonded indebtedness and timed most investments with population increase. The declining community was more dependent on intergovernmental transfers to fund public investments and the stable community timed investments so that bonded indebtedness was not required.

RALPH J. BROWN, Ph.D. Colorado 1970. A factor analysis of the differential growth process of rural-urban centers in the upper Midwest.

ALBERT M. CHURCH III, Ph.D. Claremont School 1971. An econometric model of crime in California.

A theoretical model regarding the commission and regulation of felony criminal offenses is constructed and estimated. Potential criminal offenders are assumed to maximize expected utility from legitimate and illegitimate activities. The aggregate behavior of offenders is expressed in a derived supply of offenses function. Society is hypothesized as minimizing the social costs of crime. The first and second order conditions for optimization by offenders and by society are examined. The model allows for an *ex post* adjustment to optimum crime regulation as well as possible instantaneous or stable equilibrium. This permits an empirical test of whether society acts to minimize the social costs of criminal activity.

ADNAN DAGHESTANI, Ph.D. Colorado State 1971. Analysis of regional multipliers.

An examination of estimation procedures and the conceptual framework of the input-output approach in relation to the problem of regional growth.

ROBERT D. EMERSON, Ph.D. Purdue 1971. An economic analysis of the provision of police services.

An economic model for an analysis of the provision of police services is developed. The definition of an output measure for police services is considered throughout. A measure dependent on extraneous weights for the output components is eventually used for estimation of cost equation parameters. The model permits a systematic analysis of questions of returns to scale in the provision of police services. Questions of resource allocation are approached, but statistical considerations limit their interpretation. The model is cast in such a way that questions relating to the supply of offenses are also analyzed.

WARREN E. FARB, Ph.D. Washington (St. Louis) 1971. An estimate of the relative supply and demand for substandard rental housing in major *U.S.* cities.

This cross section study constructs relative supply and demand schedules for the substandard rental housing market through the estimation of a two-equation simultaneous model. In general, the results indicate that the relative supply is probably highly elastic and that the relative demand is inelastic, and that the most efficient means of shifting the demand schedule would be through the income variable.

EDWARD J. FORD, JR., Ph.D. Boston College 1971. Benefit-cost analysis: urban renewal in the West End of Boston.

Benefit-cost analyses have often failed to place sufficient emphasis on the distributional impacts of public investments. This dissertation discusses a benefit-cost framework for residential urban renewal projects which places particular stress on distributional considerations. Utilizing this analytical framework, empirical estimates of impacts of urban renewal as it occurred in the West End of Boston are presented. These estimates indicate that distributional considerations were quite significant in the West End.

ALLEN L. FREDERICK, Ph.D. Purdue 1971. Growth and structural change in Indiana manufacturing industries, 1959–1968.

Changes in employment for the twenty-one manufacturing major industry groups in Indiana were related to national changes in employment. Approximately 56 percent of the manufacturing employees in Indiana were employed by the 100 largest firms in 1959 and 1968. Mainly due to merger activity, large firms headquartered in-state fell from 36 in 1959 to 30 in 1968, and their share of employment among the top 100 firms declined from 19.7 percent to 16.6 percent.

PAUL H. GESSAMAN, Ph.D. Cornell 1970. A study of the impact of transportation on land use and rural life in four southern tier counties of New York.

This study examined the role of transportation as a factor influencing the location of economic activity and pattern of land use in south central New York. Aerial photographs spanning the 1938–68 period and field surveys were used to reach the following conclusions: nearly all of the increased employment opportunities occurred within three miles of a major highway, and by 1968 over 85 percent of heads of rural households commuted to off-farm employment opportunities.

J. FRED GIERTZ, Ph.D. Northwestern 1970. An economic approach to the allocation of police resources.

GERALD S. GOLDSTEIN, Ph.D. Princeton 1970. Household behavior in the housing market: a cross section study of households in the San Francisco Bay area, 1965.

This thesis uses multiple linear regression techniques to study household behavior in the housing market. The study is based on a survey of households in the San Francisco Bay area in 1965. A group of households that had moved recently were studied with regard to their decisions to purchase or rent housing and with regard to the number of rooms, and the rent or house value of the housing unit selected.

Frank J. Goode, Ph.D. Minnesota 1970. An economic analysis of the survey of land for urban use.

The primary objective of this thesis was to identify the factors that affect the development of urban land, and thereby, provide city officials with information concerning the tools at their disposal for regulating urban growth. To accomplish this objective, a theoretical model of the urban land market was constructed, and hypotheses concerning the effect of various factors were deduced. A new statistical technique was used to test these hypotheses.

David M. Gordon, Ph.D. Harvard 1971. Class, productivity, and the ghetto: a study of labor market.

This dissertation presents some empirical evidence on the importance of labor market stratification as a cause of urban poverty and underemployment in the United States. The analyses build on some formal definitions of stratification, a model of the various determinant stages of labor market careers, and an original empirical definition of two job sectors—the "primary" and "secondary" markets. The analyses suggest that stratification by both demographic characteristics and by job sector constitute important causes of market inequalities.

Michael E. Granfield, Ph.D. Duke 1970. An econometric model of residential location.

Donald C. Guy, Ph.D. Illinois (Urbana) 1970. An analysis of a filtering hypothesis in urban housing markets: an empirical test in Kankakee.

Stanley W. Hamilton, Ph.D. California (Berkeley) 1970. Micro-empirical study of price movements in urban properties.

Bennett Harrison, Ph.D. Pennsylvania 1970. Education, training and the urban ghetto.

After a critique of the applicability of the theory of human capital to the explanation of nonwhite poverty, and a detailed description of the ghetto labor force, the thesis presents a micro-econometric analysis of the returns to education and training for 50,000 workers living in sixteen (individually identified) metropolitan areas in 1966. Returns are measured by individual earnings, unemployment, and occupational status. Workers are identified by intrametropolitan place of residence. The thesis utilizes unpublished *BLS* and *OEO*/Census micro data.

Michael B. Hargrove, Ph.D. Kentucky 1971. Economic development of areas contiguous to multipurpose reservoirs: the Kentucky-Tennessee experience.

This study investigates the potential and realized effect of multipurpose reservoir projects on the development of contiguous areas. Evidence of comparative levels and rates of change in income, population, and employment patterns is examined for the period 1940–1960 for twenty projects in Kentucky and Tennessee. The study concludes that planning of reservoir projects has generally been too restricted in scope to allow projects to promote rapid growth in the contiguous areas.

Andrew S. Harvey, Ph.D. Clark 1971. A reformulation of the economic base model of urban growth.

Sally M. Hey, Ph.D. Johns Hopkins 1970. A study of municipal public finances in Bogota D.E., Medellin, and Cali, Colombia.

John E. Husing, Ph.D. Claremont School 1971. The relative impact of civilian versus military defense payrolls on a small region 1964–1968.

The dissertation reports the different impacts which civilian and military payrolls had in a small region. It reviews the sectoral and multiplier impacts which occurred when the region lost a large *DOD* civilian payroll and gained a large military payroll. The model specifically allows for the effects of military consumer peculiarities such as commissary privileges and out of area allotments. The dissertation relies heavily on local economic data sources.

George D. Johnson, Ph.D. Kansas State 1970. Short-run determinants of small community growth: a multivariate analysis of factors affecting the movement of short-run mobile resources.

The ability of the small community to influence its future is expressed as a function of the economic, social, and geographic dimensions of the community base. Factor analysis and regression techniques are utilized to separate and evaluate the difference in performance between viable and nonviable communities. Communities are studied both as central cities and functionally integrated regions. The effectiveness of the social dimension is found to be closely related to the choice of growth objectives.

Jerome D. Johnson, Ph.D. Missouri 1970. The economic impact of a state university on its community.

This study investigated the economic impact of Wisconsin State University-Eau Claire on the community for the fiscal year of 1967–1968, and projected an economic impact of the state university on the community to 1975. Data was secured through a stratified random sample of faculty, classified personnel, and students, in addition to personal interviews and correspondence with university administrators and contractors.

Terry D. Kahn, Ph.D. California (Berkeley) 1970. Optimal distribution of residence locations by race: a linear programming approach.

Richard J. Kieffer, Ph.D. Missouri 1970. An economic analysis of municipal police services.

A demand and supply approach to an explanation of the per capita costs for police services in the Missouri group (sixteen Missouri nonmetropolitan municipalities with 10,000 to 50,000 population) of the municipal police industry.

Victoria C. Lapham, Ph.D. Southern Methodist 1970. Price differences for black and white housing.

This study compares prices of housing bought by blacks and whites in Dallas in 1960. Implicit values were estimated for characteristics bought by blacks and by whites. Applying these values to the average black-owned housing unit indicated the price of black housing was 4 to 6 percent less than white. However, applying these to the average white-owned unit indicated the price of white housing was 0.3 to 0.8 percent less.

Alan C. Leslie, D.B.A. Harvard 1971. An analysis of decision-making processes in Boston's municipal government.

This study describes and analyzes several decisions or decision areas of government bodies, principally those of the City of Boston. Specific topics discussed include Boston's operating and capital budget, organization of its housing agencies, the mayor's office and staff support agencies, generation of its revenues, functioning of its fiscal agencies, and creation and impact of the Home Rule Commission.

Ronald D. Lindmark, Ph.D. Ohio State 1971. Second homes in northwestern Wisconsin: a study of the owners and their use patterns and characteristics of the second home structure.

Anthony B. Low-Beer, Ph.D. Columbia 1971. The use of effluent charges to control air pollution and induce an efficient allocation of resources.

Randolph C. Martin, Ph.D. Washington (St. Louis) 1971. Differentials in the intensity of residential land use between cities and suburbs.

The objective of this study was to estimate and explain differences in the distribution of population densities over space between cities and suburbs. Net population density gradients (log-linear) were estimated for cities and suburbs of seven *U.S.* urbanized areas. These estimates show a general tendency for a smaller gradient and lower intercept for the suburban parts of the urbanized areas. The majority of this difference is explained by age of dwellings, income of inhabitants, and a price differential due to federal taxing policy and mortgage programs.

M. Amata Miller, Ph.D. California (Berkeley) 1971. Regional trends in real wages, United States, 1860–1880.

Robert H. Nelson, Ph.D. Princeton 1971. The theory of residential location.

Existing theories of residential location and rents, such as those of Alonso, Muth, and Wingo, are oversimplified. They take insufficient account of the large number of travel times and distances which determine the accessibility of any given location and of the other characteristics, such as levels of public services and neighborhood attractiveness, that determine the desirability of a location. Taking such factors into account, a theory of residential location of consumers is first outlined and then rigorously developed using analytical tools familiar from utility theory.

James C. Nicholas, Ph.D. Illinois (Urbana) 1970. The economic growth of cities: a modified export base approach.

Donald A. Nielsen, Ph.D. Syracuse 1970. Property valuation growth by county in New York State and its relation to economic development, 1955–1965.

Jonathan D. Ogur, Ph.D. Cornell 1970. The impact of colleges and universities on local renting housing markets.

Helen B. O'Neill, Ph.D. McGill 1971. Some spatial aspects of Irish economic development.

Michael Parti, Ph.D. Northwestern 1970. A factor-analytic interpretation of the educational process.

A factor analytic procedure is used to examine the underlying regularities in a set of fifty variables relevant to the educational process in urban grammar schools. The data used are a stratified sample of the data collected for the Coleman report on equality of educational opportunity. Five factors are extracted from the data. The interrelationships suggested by these factors are discussed and a causal model consistent with the factor analysis is posited.

Joseph J. Persky, Ph.D. Harvard 1971. Industralization and urban growth in the southern United States.

Much of the economic development literature implicitly assumes that industralization and urbanization are essentially elements of one process. However, in the southern United States many industries have shown a preference for rural and small-town locations. This pattern can be traced to the surplus rural labor pool of the region. The origins of this labor pool are discussed, with emphasis on the role of a "push" off the farm. A model is developed to explain the city-size distribution of manufacturing in the South as opposed to the non-South. Implications of more recent metropolitan growth for the black population of the region are explored.

John C. Pickett, Ph.D. Missouri 1970. Optimal quantities of fire department services.

The derivation of a demand curve for both fire department services and fire insurance is presented for a risk-averting property owner. The cost structure of publicly provided fire department services is discussed. The criterion of marginal benefit equals marginal cost is used to establish the optimal quantity of fire department services. A sample of thirteen cities provides the data for the statistical analysis.

JOSEPH S. RABIANSKI, Ph.D. Illinois (Urbana) 1970. Real earnings and the present value of future earnings in a theory of human migration.

CHARLES P. RAHE, Ph.D. Colorado 1970. An analysis of the role of urban industrial development agencies in the industrial location process: the metropolitan Denver experience, 1965–1970.

JAMES D. REED, Ph.D. Kansas State 1970. The measurement of the economic structure and performance of an urban economy by use of an input-output model.

The purpose of this study is to measure and analyze the economic impact on the Wichita economy resulting from changes in the level of defense expenditures in the aerospace industry. This is achieved by the construction of a thirty-nine by forty-one input-output table from 1965 survey data. Through the use of this model, traditional impact multipliers are calculated with emphasis placed on the dominant aerospace sector. While the Wichita economy is highly dependent on defense expenditures in this industry, the impact from changes in expenditures in the aerospace industry are minimized because of the weak intra-urban input-output linkages.

GEORGE R. RICE, Ph.D. Kentucky 1970. State differentials in income concentration: a simultaneous equation model explanation.

The purpose of this study is to determine what group of socioeconomic factors influence state differentials in income concentration. Theory suggests that the explanatory structure is more complex and detailed than that postulated by single equation models; therefore, a seven-equation econometric model is constructed using 1960 cross section data from the forty-eight coterminous states.

LAHOMA L. RIEDERER, Ph.D. Florida 1970. An analysis of the industrial growth of thirty-six selected standard metropolitan statistical areas from 1952 to 1966.

Employment data are utilized to analyze economic growth with three methodologies: comparative growth rates, shift analysis, and diversification. Analysis is designed to develop statistical evidence to evaluate the major hypotheses underlying the export-base and sector theories of economic growth and to determine if there are identifiable patterns that might form a basis for generalization and for projecting economic development. Analysis revealed regularities and trends that may be useful for prediction and for understanding the changes occurring.

JEFFREY M. ROMM, Ph.D. Cornell 1971. Nuclear power, Cayuga Lake, and economics.

A comprehensive examination of the physical, biological, political, and economic aspects of the proposal to construct a nuclear power plant on Cayuga Lake and the conflict situation created by groups opposing such construction.

GEORGE J. SCHIEBER, Ph.D. Syracuse 1970. A simulation study of regional disparities in per capita income.

CLAUDIA D. SCOTT, Ph.D. Duke 1971. Forecasting public outlays: an expenditure model for New Haven, Connecticut.

STEPHEN SENINGER, Ph.D. Washington (St. Louis) 1971. The regional distribution of defense procurement.

JAY SIEGEL, Ph.D. Stanford 1970. Intrametropolitan migration of white and minority group households.

A five-equation simultaneous model is estimated by 2SLS. The dependent variables are home and job location, monthly rent, number of rooms and types of dwelling unit. The data is from a transportation survey conducted in 1965 in the San Francisco Bay Area. The results support the model. For white and black households, the elasticity of home location with respect to job location is much larger in magnitude than the elasticity with respect to income.

JOSEPH F. SINGER, Ph.D. Arkansas 1971. An investigation of the role of out-migration in the socioeconomic adjustments of households left behind: a case study of the Arkansas Ozark region 1956–69.

The purpose of this study is to evaluate the influence of out-migration on the socioeconomic adjustment of households left behind in the Ozark region through a comparative analysis of family status 1956 and 1969, and to identify trends for the area as a whole.

DAVID L. SJOQUIST, Ph.D. Minnesota 1971. Property crime as an economic phenomenon.

The thesis explored the hypothesis that criminals when committing crimes against property (robbery, larceny, burglary, and auto theft), behave as rational economic beings making decisions under risk. A model, based on the von Neuman-Morgenstern hypothesis, was developed. The parameters of the resulting behavioral equation were estimated using a cross-sectional sample of the means of communities with population between 25,000 and 200,000. The empirical results lead us to accept the hypothesis.

BYUNG NAK SONG, Ph.D. Southern California 1970. A theory of regional economic growth: growth poles and development axes.

This dissertation attempts to extend Perroux's theory of growth poles and Pottier's theory of development axes in relation to existing theories of regional economic growth. It investigates the relationships between growth poles and lagging regions, and explores the use of growth poles and development axes as policy tools for regional economic development. A mathematical model is formulated and an attempt is made to test it empirically with Korean data.

Dale L. Stansbury, Ph.D. Ohio State 1970. Rural income determination in the East North Central states: implications on policy and poverty.

Norman H. Starler, Ph.D. Iowa State 1970. Planning for the optimization of resource allocation in state aid to education: an application to Iowa public schools.

In this thesis, the results of empirical investigations in three areas of the economics of education are presented. These areas are the resource redistribution effects of state aid to primary and secondary education; the educational output effects of state aid plans; and the specifications of the relation between educational output and input.

Douglas P. Sweetland, Ph.D. Kansas State 1971. Inter-SMSA migration flows: causes and consequences.

The purpose is an analysis of the determinants and consequences of inter-SMSA migration for the period 1955–1960. Part I utilizes a single equation model to analyze the determinants, using a random sample of 50 of the largest 100 SMSA's—i.e., SMSA's having a population of over 250,000 in 1960. Part II uses a simultaneous equations model on all SMSA's to extend the analysis to determine the impact of inter-SMSA migrants on the sending and receiving SMSA's.

William J. Tater, Ph.D. Southern California 1970. Impact of land reform on pacification: a quantitative appraisal of land reform in Vietnam.

Jeffry A. Timmons, D.B.A. Harvard 1971. Entrepreneurial and leadership development in an inner city ghetto and a rural depressed area.

The principal objective of the thesis was to examine the impact, relevancy and potential cost-effectiveness of entrepreneurial training as a possible vehicle for creating entrepreneurship, leadership and community development, particularly in inner city ghetto areas. The major hypothesis tested was whether or not course participants were significantly more active than their untrained counterparts.

Eric J. Toder, Ph.D. Rochester 1971. The distribution of public school teachers by race and income class in an urban metropolitan area.

In this paper, a model of market discrimination in the supply curve for public school teachers is explained and tested using data from the Boston Metropolitan Area. The available evidence indicates that school systems with more nonwhite students, face a higher supply price for equivalently qualified teachers. The resulting distribution of educational inputs is then examined along with the impact by race and income level of state and federal aid.

Howard Whitney, Ph.D. Oregon 1971. The extended school year: an economic analysis.

Kenneth Wieand, Ph.D. Washington (St. Louis) 1970. Property values and air pollution: a cross section analysis of the Saint Louis urban area.

Pollution may affect the unit price of residential housing. The output of housing per acre is an appropriate proxy for the unit price of housing. The variance of output per acre in a sample of census tracts is explained by variables measuring location, housing quality, and pollution. All exogenous variables except pollution are significantly related to output per acre.

Related Disciplines

Jan P. Acton, Ph.D. Harvard 1971. Evaluation of a life saving program: the case of heart attacks.

Ronald E. Beller, Ph.D. Florida 1970. Toward a resource planning model for community hospitals: a conceptual model of the patient treatment system.

Dinesh Bhoj, Ph.D. Pennsylvania 1971. On multivariate tests of hypotheses on mean vectors with missing observations.

John A. Byrne, D.B.A. George Washington 1970. Problems of adapting the Department of Defense cost effectiveness analysis system to nondefense agencies.

This study examines the problem of adapting the Defense Department's Cost Effectiveness Analysis (*CEA*), as an analytic subsystem of the Planning-Programming-Budgeting System, to a selected group of nondefense agencies' resource management activities. This adaptability is analyzed from the viewpoint of 1) the unique characteristics of the Defense Department's *CEA* and 2) the organizational aspects of the decision-making structure and processes, innovative environment, and innovative powers.

Michael E. Echols, Ph.D. California (Berkeley) 1970. Decision rules for problem formulation in applied management science.

Richard T. Johnson, D.B.A. Harvard 1971. Presidential style and staff behavior in the White House.

The thesis draws upon two courses of data in examining presidential style and staff behavior in the White House. The first of these sources is published material. A considerable body of literature is available describing recent presidents and their staffs. The second source of data, comprising the main body of the thesis, was acquired from a longitudinal study of the President's Council on Executive Organization. A scheme of analysis is developed to examine the relationships between the various actors in the presidential setting and the ways in which they interact.

William W. Metz, D.B.A. George Washington 1970. Identification and analysis of research and development project management problems based on nonnuclear munitions development in the Air Force.

This study evaluates research and development project management problems encountered in the Air Force and industry. Analysis of responses from each managerial group resulted in typologies for most frequently occurring and most important problems. Statistical techniques employed showed varying relationships between most important problems, time allocation, and problem categories. No positive correlation was found between authority and organizational structure. The industry manager had higher skills than the Air Force manager.

HUBERT P. MILLS, D.B.A. George Washington 1970. Management of aeronautical engineering changes (modifications) in the U.S. Navy.

The purpose of this research was to inquire into the overall management of aeronautical engineering changes (modifications) to existing weapons systems by the U.S. Navy and to ascertain their effect on the logistics support of aeronautical weapons systems in the fleet. For comparative purposes, the U.S. Air Force management system for accomplishing the Air Force aeronautical modification program was also investigated. The researcher found that while management by the Navy was satisfactory, improvements were needed. Recommendations for improvement are included in the summary chapter.

ANDREW P. MOSIER, D.B.A. George Washington 1971. An empirical analysis of the knowledge of computer technology and data processing that is useful to executives in intermediate positions in the Department of Defense.

The purpose of this exploratory study was to begin the formation of an empirical base for the development of computer-based information systems curriculum for executives. The research question focused on an underlying field—computer technology and data processing (*CT&DP*). It sought to learn the relative usefulness of knowledge of nineteen *CT&DP* subtopics to executives in intermediate positions in large complex organizations.

SAMUEL L. NEWMAN, Ph.D. Alabama 1970. The use of opinion technology in long-range planning: a study of the Delphi process in forecasting United Fund needs in Tuscaloosa, Alabama.

DENNIS W. ORGAN, Ph.D. North Carolina 1970. Some factors influencing the behavior of boundary role persons.

A lab experiment was conducted to test the effects of three variables on the behavior of persons engaged in intergroup conflict resolution. The variables were: 1) the perceived behavioral norm of the bargainer's constituency; 2) the degree to which the bargainers could be monitored by constituents; and 3) the degree of confidence expressed in the bargainers by constituents. Variable 1) had the strongest effect, variable 2) a smaller but significant effect, and variable 3) an insignificant effect.

CHARLES J. PEARCE, D.B.A. George Washington 1970. An appraisal of hospital administration education in the U.S. Navy.

This study reports on the attitudes and opinions of senior Naval hospital administrators and recent graduates of the Naval School of Hospital Administration with respect to the curriculum of its undergraduate course. It seeks to provide Navy officials and educators with basic knowledge that may be useful in designing future curricula. A secondary benefit is to provide military and civilian hospital administration schools with information on undergraduate hospital administration curricula.

GERALD PERSELAY, D.B.A. George Washington 1970. A study of U.S. Air Force experience under the Federal Employee-Management Cooperation Program.

This study examines Air Force experience under Executive Order 10988. Written agreements between Air Force bases and union locals are analyzed. Problems in the program are outlined by government and union leaders. Local union and management problems are sampled using an attitudinal mail survey and results are statistically analyzed. Conclusions are drawn and recommendations made for an improved program.

PETER B. PETERSEN, D.B.A. George Washington 1971. An investigation of the effect of training.

The purpose of the study was to investigate the effect of U.S. Army Engineer Officer Candidate training and the stability of training when measured again by the *JAIM* instrument three years after departure. Significant differences were found for fourteen of fifteen predicted scales pertaining to the effect of training. With respect to stability of training, four of twelve scales were significant in the predicted direction, three in opposite direction, and no significant difference for five other scales.

KARL F. PRICE, Ph.D. Pennsylvania 1970. A study of role conflict in the relationship between the art museum director and his board of trustees.

The study is divided into two portions: the development of a theoretical framework to study role conflict, and the gathering of empirical data, from art museums, to test hypotheses generated from the model. In the first portion of the study, a taxonomy of role conflict was developed. From the data gathered through interviews at fourteen museums, operational measures of role conflict were developed and the hypotheses related to the model were tested.

BERNARD SISKEN, Ph.D. Pennsylvania 1970. Reserve decision discrimination.

ROBERT M. SPRINGER, JR., D.B.A. George Washington 1970. An analysis of attitudes toward the legitimacy of supervisory influence in a military environment.

This study hypothesized that, in the Army organization tested, there would be differences between the attitudes of first line supervisors and their subordinates toward the legitimacy of organizational influence, and that these differences would be associated with certain forms of undesirable group characteristics. The results of statistical testing of 1,200 Army basic trainees supported the hypothesis that differences in attitudes exist, but it was not possible to associate these differences with undesirable group characteristics.

Thomas Tirney, Ph.D. Pennsylvania 1971. A study conformity behavior.

Small experimental decision-making groups in a face-to-face situation were used to generate the data used to test hypotheses concerning the group variables of attractiveness of a group to a member, monetary reward received by a member because of his membership in th group, and the difference between a member's opinio and the group's opinions.